
WORLD (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Tue Dec 22 05:19:34 1998; MasPar time 285.89 Seconds
Tubular output not generated. 1077.389 Million cell updates/sec

Title: >US-08-530-112A-2965
Description: (1-172) from US08530112A.seq
Perfect Score: 171
N.A. Sequence: 171
Comp: 1 GATCTGACGATGCCGTGTC.....AAATAAATAAATAGTGAAA 172
CTAGACTGCTAGGCACAG.....TTTATTATTATTCACCTTT

Scoring table:
Gap 6
TABLE default

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-est55
Database: 1:em_est1 2:em_gss1 3:em_gss2 4:em_gss3
genbank-est107
5:gb_est1 6:gb_est10 7:gb_est11 8:gb_est12 9:gb_est13
10:gb_est14 11:gb_est15 12:gb_est16 13:gb_est17
14:gb_est18 15:gb_est19 16:gb_est2 17:gb_est20
18:gb_est21 19:gb_est3 20:gb_est4 21:gb_est5 22:gb_est6
23:gb_est7 24:gb_est8 25:gb_est9 26:gb_gss1 27:gb_gss2
28:gb_gss3 29:gb_gss4

Statistics: Mean 9.562; Variance 2.070; scale 4.620

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
C 1	171	100.0	281	17	AI004135	ot43g02.s1 Soares_test 4.15e-265
C 2	171	100.0	344	12	AA724792	ai05e02.s1 Soares para 4.15e-265
C 3	171	100.0	341	10	AA626315	zu93e06.s1 Soares test 4.15e-265
C 4	171	100.0	394	7	AA448861	zx11d01.s1 Soares tota 4.15e-265
C 5	171	100.0	420	18	AI077370	oy87f05.x1 Soares_feta 4.15e-265
C 6	171	100.0	423	20	N78952	zb18c05.s1 Homo sapien 4.15e-265
C 7	171	100.0	438	10	AA435896	zt80e12.s1 Soares test 4.15e-265
C 8	171	100.0	437	21	W68342	zd34c01.s1 Soares_feta 4.15e-265
C 9	171	100.0	447	18	AI147031	ga82c01.x1 Soares_feta 4.15e-265
C 10	171	100.0	449	18	AI092949	qb11c06.x1 Soares_feta 4.15e-265
C 11	171	100.0	468	20	N78829	zb17a05.s1 Homo sapien 4.15e-265
C 12	171	100.0	549	13	AA824370	aj29a06.s1 Soares test 4.15e-265
C 13	169	98.8	405	15	AA991266	os31c03.s1 NCI_CGAP_Br 2.57e-261

C 14	168	98.2	379	13	AA862620	oh07c11.s1 NCI_CGAP_K1 2.02e-259
C 15	168	98.2	407	5	T98518	ye60r05.s1 Homo sapien 2.02e-259
C 16	168	98.2	450	17	AI015595	ov09h08.x1 NCI_CGAP_K1 2.02e-259
C 17	165	96.5	481	22	W93038	z973b04.s1 Soares_feta 9.72e-254
C 18	162	94.7	520	22	AA027214	ze97b09.s1 Soares_feta 4.61e-248
C 19	160	93.6	481	20	W17346	zb18c05.r1 Soares_feta 2.78e-244
C 20	151	88.3	470	7	AA447925	zx11d01.r1 Soares tota 2.62e-227
C 21	146	85.4	253	10	AA652731	ns65b03.s1 NCI_CGAP_Pr 6.60e-218
C 22	135	78.9	477	20	W17259	zt57g10.r1 Soares_feta 2.66e-197
C 23	113	66.1	424	8	AA401790	zd34c01.r1 Soares test 1.73e-156
C 24	110	64.3	442	21	W68383	qj22c11.s1 NCI_CGAP_K1 5.66e-151
C 25	80	46.8	262	14	AA878662	cj02a03.s1 Soares_feta 1.14e-89
C 26	76	44.4	388	18	AI143531	ye98b02.r1 Homo sapien 5.76e-74
C 27	67	39.2	334	5	R07662	UI-R-C2p-ok-d-10-0-UI. 1.11e-63
C 28	61	35.7	441	18	AI137425	97SN1784 Rice Immature 2.98e-37
C 29	45	26.3	247	12	AA754458	zm53a01.r1 Stratagene 2.98e-37
C 30	45	26.3	504	11	AA070633	2b72d07.r1 Soares_feta 5.83e-31
C 31	41	24.0	326	20	W25674	97SN1787 Rice Immature 7.59e-25
C 32	37	21.6	252	12	AA754459	97SN1787 Rice Immature 2.38e-23
C 33	36	21.1	352	5	R07618	ye98b02.s1 Homo sapien 6.00e-19
C 34	33	19.3	2275	11	AF034173	Homo sapiens ncon2 co 2.51e-09
C 35	28	16.4	247	12	AA754458	C.elegans cDNA clone Y 1.42e-05
C 36	26	15.2	2275	11	AF034173	HS_3126_B2_B11_MR Cit 1.42e-05
C 37	23	13.5	300	9	C57680	15720.seq.F.Fetal hear 1.42e-05
C 38	23	13.5	336	22	AA095731	Homo sapiens ncon2 co 1.42e-05
C 39	23	13.5	362	28	AQ187317	HS-1025-B2-E03-MF.abi 2.18e-04
C 40	22	12.9	281	26	B34858	mb95d08.r1 Soares_mous 2.18e-04
C 41	22	12.9	415	24	AA267233	vz88c09.r1 Soares_mous 2.18e-04
C 42	22	12.9	475	24	AA285599	vj80a06.r1 Soares_mous 2.18e-04
C 43	22	12.9	588	8	AA542181	RPC111-58M17.TU RPC111 2.18e-04
C 44	22	12.9	585	28	AQ201588	
C 45	22	12.9				

ALIGNMENTS

RESULT 1	AI004135	281 bp	mRNA	EST	27-AUG-1998
LOCUS	ot43g02.s1	Soares_testis_NHT	Homo sapiens	cDNA clone	IMAGE:1619570
DEFINITION	3', mRNA sequence.				
ACCESSION	AI004135				
NID	Q3213645				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 281)				
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
COMMENT	Unpublished (1997)				

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 583 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 152.
Location/Qualifiers
1. .281

/organism="Homo sapiens"
/note="Vector: p7T3p-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech

Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5',
 TGTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo." /db_xref="taxon:9606"
 /clone="IMAGE:1619570"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /lab_host="DH10B (ampicillin resistant)"

BASE COUNT 69 a 67 c 52 g 93 t
 ORIGIN

Query Match 100.0%; Score 171; DB 17; Length 281;
 Best Local Similarity 99.4%; Pred. No. 4.15e-265;
 Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 28 TTTCACCTTATTTATTTATTTTCCCAATTTCTCTTCCAGACTCCATCCCAAGAGT 87
 Cp 172 TTTCACCTTATTTATTTATTTTCCCAATTTCTCTTCCAGACTCCATCCCAAGAGT 113
 Db 88 CATACAGCCTTCTCCACCTTCTACATGAATATACATCCCACTGGAACAAGGCACAG 147
 Cp 112 NATAGCAGCCTTCTCCACCTTCTACATGAATATACATCCCACTGGAACAAGGCACAG 53
 Db 148 ACAGGAGGAGGGGATAGGACTTCGCAACTGGACAGGCGATCGTTGAGATC 199
 Cp 52 ACAGGAGGAGGGGATAGGACTTCGCAACTGGACAGGCGATCGTTGAGATC 1

RESULT 2 AA724792 314 bp mRNA EST 08-JAN-1998
 LOCUS ai05e02.s1 Soares parathyroid tumor NHPA Homo sapiens cDNA clone
 DEFINITION 1341914 3', mRNA sequence.
 ACCESSION AA724792
 NID G2742499
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 314)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 244.
 Location/Qualifiers

FEATURES
 source
 1. 314
 /organism="Homo sapiens"
 /note="Organ: parathyroid gland; Vector: pT73D (Pharmacia) with a modified polylinker; Site1: Not I; Site2: Eco RI; oligo(dT) primer [5', TGTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to

Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."

/db_xref="taxon:9606"
 /clone="1341914"
 /clone_lib="Soares parathyroid tumor NHPA"
 /tissue_type="parathyroid tumor"
 /dev_stage="adult"
 /lab_host="DH10B (ampicillin resistant)"

BASE COUNT 72 a 80 c 58 g 104 t
 ORIGIN

Query Match 100.0%; Score 171; DB 12; Length 314;
 Best Local Similarity 99.4%; Pred. No. 4.15e-265;
 Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 11 TTTCACCTTATTTATTTATTTTCCCAATTTCTCTTCCAGACTCCATCCCAAGAGT 70
 Cp 172 TTTCACCTTATTTATTTATTTTCCCAATTTCTCTTCCAGACTCCATCCCAAGAGT 113
 Db 71 CATAGCAGCCTTCTCCACCTTCTACATGAATATACATCCCACTGGAACAAGGCACAG 130
 Cp 112 NATAGCAGCCTTCTCCACCTTCTACATGAATATACATCCCACTGGAACAAGGCACAG 53
 Db 131 ACAGGAGGAGGGGATAGGACTTCGCAACTGGACAGGCGATCGTTGAGATC 182
 Cp 52 ACAGGAGGAGGGGATAGGACTTCGCAACTGGACAGGCGATCGTTGAGATC 1

RESULT 3 AA626315 341 bp mRNA EST 15-OCT-1997
 LOCUS zu99e06.s1 Soares testis NHT Homo sapiens cDNA clone 745570 3',
 DEFINITION mRNA sequence.
 ACCESSION AA626315
 NID G2538702
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 341)
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
 WashU-NCI human EST Project
 Unpublished (1997)

TITLE
 JOURNAL
 COMMENT

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES
 source
 1. 341
 /organism="Homo sapiens"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site1: Not I; Site2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

/db_xref="GDB:593352"
/db_xref="taxon:9606"
/clone="745570"
/clone_lib="Soares testis NHT"
/sex="male"
/lab_host="DH10B"
/db_xref="GDB:5983764"

BASE COUNT 84 a 85 c 65 g 107 t

Query Match 100.0%; Score 171; DB 10; Length 341;
Best Local Similarity 99.4%; Pred. No. 4.15e-265;
Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 5 TTTCACTTATTATTTATTTTATCTTCCAAATTCCTCTGCGAGACTCCCATCAAAGAGT 64

|||||

Cp 172 TTTCACTTATTATTTATTTTATCTTCCAAATTCCTCTGCGAGACTCCCATCAAAGAGT 113

|||||

Db 65 CATACGAGCTTCTCCACCTTCTACATGAATACATCCCACTGACCAAGGCACACG 124

|||||

Cp 112 NATAGCAGCTTCTCCACCTTCTACATGAATACATCCCACTGACCAAGGCACACG 53

|||||

Db 125 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACAGCGCATCGTTCAGATC 176

|||||

Cp 52 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACAGCGCATCGTTCAGATC 1

|||||

RESULT 4 AA448861 394 bp mRNA EST 04-JUN-1997

LOCUS zxl1d01.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone

DEFINITION 786145 3', mRNA sequence.

ACCESSION AA448861

NID g2162531

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;

Homo.

1 (bases 1 to 394)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,

Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,

Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,

White,X., Wylie,T., Waterston,R. and Willson,R.

WashU-Merck EST Project 1997

Unpublished (1997)

CONTACT: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -41m13 fwd. ET from Amersham.

Location/Qualifiers

1..394

/organism="Homo sapiens"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was prepared from mRNA obtained from pooled 8-9 week

(total) fetus material with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAGTGGGACGCGCTTAATTTTATTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

/db_xref="taxon:9606"

BASE COUNT 84 a 85 c 65 g 107 t

Query Match 100.0%; Score 171; DB 10; Length 341;
Best Local Similarity 99.4%; Pred. No. 4.15e-265;
Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 5 TTTCACTTATTATTTATTTTATCTTCCAAATTCCTCTGCGAGACTCCCATCAAAGAGT 64

|||||

Cp 172 TTTCACTTATTATTTATTTTATCTTCCAAATTCCTCTGCGAGACTCCCATCAAAGAGT 113

|||||

Db 65 CATACGAGCTTCTCCACCTTCTACATGAATACATCCCACTGACCAAGGCACACG 124

|||||

Cp 112 NATAGCAGCTTCTCCACCTTCTACATGAATACATCCCACTGACCAAGGCACACG 53

|||||

Db 125 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACAGCGCATCGTTCAGATC 176

|||||

Cp 52 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACAGCGCATCGTTCAGATC 1

|||||

RESULT 4 AA448861 394 bp mRNA EST 04-JUN-1997

LOCUS zxl1d01.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone

DEFINITION 786145 3', mRNA sequence.

ACCESSION AA448861

NID g2162531

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;

Homo.

1 (bases 1 to 394)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,

Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,

Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,

White,X., Wylie,T., Waterston,R. and Willson,R.

WashU-Merck EST Project 1997

Unpublished (1997)

CONTACT: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -41m13 fwd. ET from Amersham.

Location/Qualifiers

1..394

/organism="Homo sapiens"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was prepared from mRNA obtained from pooled 8-9 week

(total) fetus material with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAGTGGGACGCGCTTAATTTTATTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

/db_xref="taxon:9606"

BASE COUNT 84 a 85 c 65 g 107 t

Query Match 100.0%; Score 171; DB 10; Length 341;
Best Local Similarity 99.4%; Pred. No. 4.15e-265;
Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 5 TTTCACTTATTATTTATTTTATCTTCCAAATTCCTCTGCGAGACTCCCATCAAAGAGT 64

|||||

Cp 172 TTTCACTTATTATTTATTTTATCTTCCAAATTCCTCTGCGAGACTCCCATCAAAGAGT 113

|||||

Db 65 CATACGAGCTTCTCCACCTTCTACATGAATACATCCCACTGACCAAGGCACACG 124

|||||

Cp 112 NATAGCAGCTTCTCCACCTTCTACATGAATACATCCCACTGACCAAGGCACACG 53

|||||

Db 125 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACAGCGCATCGTTCAGATC 176

|||||

Cp 52 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACAGCGCATCGTTCAGATC 1

|||||

RESULT 4 AA448861 394 bp mRNA EST 04-JUN-1997

LOCUS zxl1d01.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone

DEFINITION 786145 3', mRNA sequence.

ACCESSION AA448861

NID g2162531

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;

Homo.

1 (bases 1 to 394)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,

Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,

Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,

White,X., Wylie,T., Waterston,R. and Willson,R.

WashU-Merck EST Project 1997

Unpublished (1997)

CONTACT: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -41m13 fwd. ET from Amersham.

Location/Qualifiers

1..394

/organism="Homo sapiens"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was prepared from mRNA obtained from pooled 8-9 week

(total) fetus material with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAGTGGGACGCGCTTAATTTTATTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

/db_xref="taxon:9606"

BASE COUNT 84 a 85 c 65 g 107 t

Query Match 100.0%; Score 171; DB 10; Length 341;
Best Local Similarity 99.4%; Pred. No. 4.15e-265;
Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 5 TTTCACTTATTATTTATTTTATCTTCCAAATTCCTCTGCGAGACTCCCATCAAAGAGT 64

|||||

Cp 172 TTTCACTTATTATTTATTTTATCTTCCAAATTCCTCTGCGAGACTCCCATCAAAGAGT 113

|||||

Db 65 CATACGAGCTTCTCCACCTTCTACATGAATACATCCCACTGACCAAGGCACACG 124

|||||

Cp 112 NATAGCAGCTTCTCCACCTTCTACATGAATACATCCCACTGACCAAGGCACACG 53

|||||

Db 125 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACAGCGCATCGTTCAGATC 176

|||||

Cp 52 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACAGCGCATCGTTCAGATC 1

|||||

RESULT 4 AA448861 394 bp mRNA EST 04-JUN-1997

LOCUS zxl1d01.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone

DEFINITION 786145 3', mRNA sequence.

ACCESSION AA448861

NID g2162531

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;

Homo.

1 (bases 1 to 394)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,

Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,

Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,

White,X., Wylie,T., Waterston,R. and Willson,R.

WashU-Merck EST Project 1997

Unpublished (1997)

CONTACT: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -41m13 fwd. ET from Amersham.

Location/Qualifiers

1..394

/organism="Homo sapiens"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was prepared from mRNA obtained from pooled 8-9 week

(total) fetus material with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAGTGGGACGCGCTTAATTTTATTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

Query Match 100.0%; Score 171; DB 18; Length 420;
 Best Local Similarity 99.4%; Pred. No. 4.15e-265;
 Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 2 TTTCACCTTATTTATTTATCTTCCAAATTCCTCTTGCAGACTCCCATCCAAAGAGT 61
 |||||
 Cp 172 TTTCACCTTATTTATTTATCTTCCAAATTCCTCTTGCAGACTCCCATCCAAAGAGT 113
 |||||

Db 62 CATAGCAGCCTTCTTCCACCTTCTACATGAATACATCCCACTGAACAAAGGCACAG 121
 |||||
 Cp 112 NATAGCAGCCTTCTTCCACCTTCTACATGAATACATCCCACTGAACAAAGGCACAG 53
 |||||

Db 122 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACACGGCATGTCAGATC 173
 |||||
 Cp 52 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACACGGCATGTCAGATC 1

RESULT 6 N78952 423 bp mRNA EST 29-MAR-1996
 LOCUS zbl8c05.s1 Homo sapiens cDNA clone 302408 3'
 DEFINITION N78952
 ACCESSION g1241653
 NID
 KEYWORDS EST.
 SOURCE human clone-302408 primer-m13 -40 forward library-Soares fetal lung
 NHLL19W vector-pT73D (Pharmacia) with a modified polylinker
 host-DH10B (ampicillin resistant) Rsite1-Not I Rsite2-Eco RI 19
 week fetus. 1st strand cDNA was primed with a Not I - oligo(dT)
 primer [5'-TGTACCAATCTCAAGTGGAGCGCGCATTTTTTTTTTTT-3'],
 double-stranded cDNA was size selected, ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I and Eco
 RI sites of a modified pT73 vector (Pharmacia). Library went
 through one round of normalization to a Cot = 5. Library
 constructed by Bento Soares and M. Fatima Bonaldo.
 Homo sapiens
 ORGANISM
 Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
 Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
 Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
 Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 423)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT
 Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 High quality sequence stops: 255
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 FEATURES
 source 1..423
 /organism="Homo sapiens"
 /clone="302408"
 BASE COUNT 107 a 105 c 77 g 133 t 1 others
 ORIGIN
 <1..>423

Query Match 100.0%; Score 171; DB 20; Length 423;
 Best Local Similarity 99.4%; Pred. No. 4.15e-265;
 Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 3 TTTCACCTTATTTATTTATCTTCCAAATTCCTCTTGCAGACTCCCATCCAAAGAGT 62
 |||||

Cp 172 TTTCACCTTATTTATTTATCTTCCAAATTCCTCTTGCAGACTCCCATCCAAAGAGT 113
 |||||

Db 63 CATAGCAGCCTTCTTCCACCTTCTACATGAATACATCCCACTGAACAAAGGCACAG 122
 |||||

Cp 112 NATAGCAGCCTTCTTCCACCTTCTACATGAATACATCCCACTGAACAAAGGCACAG 53
 |||||

Db 123 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACACGGCATGTCAGATC 174
 |||||

Cp 52 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACACGGCATGTCAGATC 1

RESULT 7 AA435896 428 bp mRNA EST 09-NOV-1997
 LOCUS zt80e12.s1 Soares testis NHT Homo sapiens cDNA clone 728686 3'
 DEFINITION mRNA sequence.
 ACCESSION AA435896
 NID g2140810
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 428)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
 Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
 TITLE WashU-NCI human EST Project
 JOURNAL Unpublished (1997)
 COMMENT
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 698 Std Error: 0.00
 Seq primer: -41m13 fwd. ET from Amersham
 High quality sequence stop: 392.
 Location/Qualifiers
 1..428
 /organism="Homo sapiens"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech
 Laboratories, Inc., and primed with a Not I - oligo(dT)
 primer [5',
 TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization to Cot5 and was
 constructed by Bento Soares and M. Fatima Bonaldo.
 /db_xref="GDB:5925599"
 /db_xref="taxon:9606"
 /clone="728686"
 /sex="male"
 /sex="male"
 /lab_host="DH10B"
 BASE COUNT 106 a 104 c 77 g 141 t
 ORIGIN
 Query Match 100.0%; Score 171; DB 10; Length 428;
 Best Local Similarity 99.4%; Pred. No. 4.15e-265;
 Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 14 TTTCACCTTATTTATTTATCTTCCAAATTCCTCTTGCAGACTCCCATCCAAAGAGT 73
 |||||

Cp 172 TTTCACCTTATTTATTTATCTTCCAAATTCCTCTTGCAGACTCCCATCCAAAGAGT 113
 |||||

Db 74 CATAGCAGCCTTCTTCCACCTTCTACATGAATACATCCCACTGAACAAAGGCACAG 133
 |||||

Cp 112 NATAGCAGCCTTCTCCACCTTCTACATGAATACATCCCCACCTGAAACAAGGCACAG 53
 Db 134 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTTCAGATC 185
 Cp 52 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTTCAGATC 1

RESULT 8 W68342 437 bp mRNA EST 15-OCT-1996
 LOCUS zd34c01.s1 Soares fetal heart NBHH19W Homo sapiens cDNA clone
 DEFINITION 342528 3', mRNA sequence.
 ACCESSION W68342
 NID g1377212
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 437)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Mafrá, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevasakis, E., Waterston, R., Williamson, A., Woldmann, P. and
 Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT

Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 521 Std Error: 0.00
 Seq primer: mob.REGA+ET
 High quality sequence stop: 254.

FEATURES
 source
 1..437
 /organism="Homo sapiens"
 /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
 modified polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT7T3 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by
 M.Fatima Bernaldo. This library was constructed from the
 same fetus as the fetal lung library. Soares fetal lung
 NBHL19W."
 /db_xref="taxon:9606"
 /clone="342528"
 /clone_lib="Soares fetal heart NBHH19W"
 /sex="unknown"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 complement(<1..>437)
 mRNA 108 a 110 c 83 g 134 t 2 others
 BASE COUNT 108 a 110 c 83 g 134 t 2 others
 ORIGIN

Query Match 100.0%; Score 171; DB 21; Length 437;
 Best Local Similarity 99.4%; Pred. No. 4.15e-265;
 Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 2 TTTCACCTATTATTATTATTTTCTCCAAATTCCTCTTGCAGACTCCCAACAAAGAGT 61
 Cp 172 TTTCACCTATTATTATTATTTTCTCCAAATTCCTCTTGCAGACTCCCAACAAAGAGT 113

Db 62 CATAGCAGCCTTCTCCACCTTCTACATGAATACATCCCACTGAAACAAGGCACAG 121
 Cp 112 NATAGCAGCCTTCTCCACCTTCTACATGAATACATCCCACTGAAACAAGGCACAG 53
 Db 122 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTTCAGATC 173
 Cp 52 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTTCAGATC 1

RESULT 9 A1147031 447 bp mRNA EST 29-SEP-1998
 LOCUS ga82c01.x1 Soares fetal heart NBHH19W Homo sapiens cDNA clone
 DEFINITION IMAGE:1693248 3', mRNA sequence.
 ACCESSION A1147031
 NID g3674713
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 447)
 AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40ml3 fwd. ET from Amersham.
 Location/Qualifiers
 1..447
 /organism="Homo sapiens"
 /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
 modified polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT7T3 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by
 M.Fatima Bernaldo. This library was constructed from the
 same fetus as the fetal lung library. Soares fetal lung
 NBHL19W."
 /db_xref="taxon:9606"
 /clone="IMAGE:1693248"
 /clone_lib="Soares fetal heart NBHH19W"
 /sex="unknown"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 BASE COUNT 114 a 112 c 80 g 141 t
 ORIGIN

Query Match 100.0%; Score 171; DB 18; Length 447;
 Best Local Similarity 99.4%; Pred. No. 4.15e-265;
 Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 4 TTTCACCTATTATTATTATTTTCTCCAAATTCCTCTTGCAGACTCCCAACAAAGT 63
 Cp 172 TTTCACCTATTATTATTATTTTCTCCAAATTCCTCTTGCAGACTCCCAACAAAGT 113
 Db 64 CATAGCAGCCTTCTCCACCTTCTACATGAATACATCCCACTGAAACAAGGCACAG 123
 Cp 112 NATAGCAGCCTTCTCCACCTTCTACATGAATACATCCCACTGAAACAAGGCACAG 53
 Db 124 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTTCAGATC 175
 Cp 52 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTTCAGATC 1

```

RESULT 10
LOCUS   AT092949      449 bp      mRNA      EST      05-OCT-1998
DEFINITION   q81c06.x1 Soares fetal heart_NBHH19W Homo sapiens cDNA clone
              IMAGE:1693162 3', mRNA sequence.
ACCESSION   AT092949
NID        93431925
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 449)
            NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)

COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 525 Std Error: 0.00
Seq primer: -40ml3 fwd. Er from Amersham.
            Location/Qualifiers
              1..449
                /organism="Homo sapiens"
                /note="Organ: heart; Vector: pT73D (Pharmacia) with a
                modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                strand cDNA was primed with a Not I - oligo(dT) primer [5',
                TGTACCAATCTGAAGTGGAGCGCGCATCTTTTTTTTTTTT 3'],
                double-stranded cDNA was size selected, ligated to Eco RI
                adapters (Pharmacia), digested with Not I and cloned into
                the Not I and Eco RI sites of a modified pT73 vector
                (Pharmacia). Library went through one round of
                normalization to a Cot = 5. Library constructed by
                M.Patima Bonaldo. This library was constructed from the
                same fetus as the fetal lung library, Soares fetal lung
                NBHL19W."
                /db_xref="taxon:9606"
                /clone="IMAGE:1693182"
                /clone_lib="Soares_fetal_heart_NBHH19W"
                /sex="unknown"
                /dev_stage="19 weeks"
                /lab_host="DH10B (ampicillin resistant)"

BASE COUNT      114 a 112 c 82 g 141 t
ORIGIN

Query Match      100.0%; Score 171; DB 18; Length 449;
Best Local Similarity 99.4%; Pred. No. 4.15e-265;
Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 4 TTTCACTATTATTATTATTATCTCCAAATTTCTCTTGGCCAGACTCCCATCCAAAGAGT 63
|||||
Cp 172 TTTCACTATTATTATTATTATCTTCCAAATTTCTCTTGGCCAGACTCCCATCCAAAGAGT 113
|||||

Db 64 CATAGCAGCTTCTTCACCTTCTACATGAATACATCCCCACCTGACACGACGACG 123
|||||
Cp 112 NATAGCAGCTTCTTCACCTTCTACATGAATACATCCCCACCTGACACGACGACG 53
|||||

Db 124 ACAGGAGGGGGAATAGGACTTCGCAAACTGGACAGGCATCGTTCAGATC 175
|||||
Cp 52 ACAGGAGGGGGAATAGGACTTCGCAAACTGGACAGGCATCGTTCAGATC 1

RESULT 11
LOCUS   N78829      468 bp      mRNA      EST      29-MAR-1996
DEFINITION   zbl7a05.s1 Homo sapiens cDNA clone 302288 3'.
ACCESSION   N78829
NID        G1241530
KEYWORDS    EST.
SOURCE      human clone-302288 primer=ml3 -40 forward library=Soares fetal lung
            NBHL19W vector=pT73D (Pharmacia) with a modified polylinker

```

JOURNAL
COMMENT

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bernaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 465.

FEATURES
source

1..549
Location/Qualifiers
/organism="Homo sapiens"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5].
TGTTCACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bernaldo."
/db_xref="taxon:9606"
/clone_lib="1391698"
/clone_lib="Soares testis NHT"
/sex="male"
/lab_host="DH10B"

BASE COUNT 139 a 140 c 101 g 169 t

Query Match 100.0%; Score 171; DB 13; Length 549;

Best Local Similarity 99.4%; Pred. No. 4.15e-265;

Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 6 TTTCACCTATTATTATTTATCTTCCAAATTCCTCTTGGCAGACTCCCATCCAAAGAGT 65

|||||

Cp 172 TTTCACCTATTATTATTTATCTTCCAAATTCCTCTTGGCAGACTCCCATCCAAAGAGT 113

|||||

Db 56 CATAGCAGCTTCTTCCACCTTCTACATGAATACATCCCACTGACAAAGGCACAG 125

|||||

Cp 112 NATAGCAGCTTCTTCCACCTTCTACATGAATACATCCCACTGACAAAGGCACAG 53

|||||

Db 126 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACAGCGCATCGTTCAGATC 177

|||||

Cp 52 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACAGCGCATCGTTCAGATC 1

RESULT 13 AA991266 405 bp mRNA EST 27-JUL-1998

LOCUS os31c03.s1 NCI_CGAP_Br2 Homo sapiens CDNA clone IMAGE:1606948 3',

DEFINITION mRNA sequence.

ACCESSION AA991266

NID 9317755

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 405)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 572 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 386.

FEATURES
source

1..405
Location/Qualifiers
/organism="Homo sapiens"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled bulk breast tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. This library is the normalized version of NCI_CGAP_Br1.1. Library was constructed by Bento Soares and M. Fatima Bernaldo."
/db_xref="taxon:9606"
/clone_lib="1606948"
/clone_lib="NCI_CGAP_Br2"
/sex="female, pooled"
/tissue_type="breast"
/lab_host="DH10B"

BASE COUNT 107 a 76 c 126 t

Query Match 98.8%; Score 169; DB 15; Length 405;

Best Local Similarity 98.8%; Pred. No. 2.57e-261;

Matches 170; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 21 TTTCACCTATTATTATTTATCTTCCAAATTCCTCTTGGCAGACTCCCATCCAAAGAGT 80

|||||

Cp 172 TTTCACCTATTATTATTTATCTTCCAAATTCCTCTTGGCAGACTCCCATCCAAAGAGT 113

|||||

Db 81 CATAGCAGCTTCTTCCACCTTCTACATGAATACATCCCACTGACAAAGGCACAG 140

|||||

Cp 112 NATAGCAGCTTCTTCCACCTTCTACATGAATACATCCCACTGACAAAGGCACAG 53

|||||

Db 141 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACAGCGCATCGTTCAGATC 192

|||||

Cp 52 ACAGGAGGAGGGAATAGGACTTCGCAAACTGGACAGCGCATCGTTCAGATC 1

RESULT 14 AA862620 379 bp mRNA EST 11-MAR-1998

LOCUS oh07c11.s1 NCI_CGAP_Kid3 Homo sapiens CDNA clone IMAGE:1457108 3',

DEFINITION mRNA sequence.

ACCESSION AA862620

NID 92955099

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 379)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome
Clone distribution: NCI-CCAP clone distribution
Information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 325.
Location/Qualifiers
1. .379

```

/organism="Homo sapiens"
/note="Organ: kidney; Vector: p7T3P-pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified p7T3 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="IMAGE:1457108"
/clone_lib="NCI_CCAP_Kid3"
/lab_host="DH10B"
99 a 92 c 75 g 113 t

```

```

BASE COUNT
ORIGIN
99 a 92 c 75 g 113 t
/lab_host="DH10B"

```

Query Match 98.2%; Score 168; DB 13; Length 379;
Best Local Similarity 99.4%; Pred. No. 2.02e-259;
Matches 168; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

[illegible]

RESULT	15
LOCUS	T98518
DEFINITION	407 bp mRNA
ACCESSION	Y660R05.s1 Homo sapiens CDNA clone 122169 3'
NID	g748255
KEYWORDS	EST.
	31-MAR-1995

human clone-122169 library-Soares fetal liver spleen INFLS vector-pf773D (Pharmacia) with a modified polylinker host-DH10B (ampicillin resistant) primer-21ml3 Rsite1-Pac I Rsite2-Eco RI Liver and spleen from a 20 week post conception male fetus. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5'-AACTGGAGAATAATTAAGAGCTTTTCTTTTCTTTTCTTTT 3'], cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pW7 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM

Homo sapiens
 Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 / bases 1,50,493
 REFERENCE

REFERENCE

cases 1 to 407).
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevisakis, E., Waterston, R., Williamson, A., Wohltdam, P. and
 Wilson, R.
 AUTHORS

TITLE	COMMENT
Wilson, R. The WashU-Merck EST Project Unpublished (1995)	

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 241
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Location: image.llnl.gov

[illegible]

BASE COUNT	ORIGIN	97 a	99 c	98 g	110 t	3 others
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	5	5
6	6	6	6	6	6	6
7	7	7	7	7	7	7
8	8	8	8	8	8	8
9	9	9	9	9	9	9
10	10	10	10	10	10	10
11	11	11	11	11	11	11
12	12	12	12	12	12	12
13	13	13	13	13	13	13
14	14	14	14	14	14	14
15	15	15	15	15	15	15
16	16	16	16	16	16	16
17	17	17	17	17	17	17
18	18	18	18	18	18	18
19	19	19	19	19	19	19
20	20	20	20	20	20	20
21	21	21	21	21	21	21
22	22	22	22	22	22	22
23	23	23	23	23	23	23
24	24	24	24	24	24	24
25	25	25	25	25	25	25
26	26	26	26	26	26	26
27	27	27	27	27	27	27
28	28	28	28	28	28	28
29	29	29	29	29	29	29
30	30	30	30	30	30	30
31	31	31	31	31	31	31
32	32	32	32	32	32	32
33	33	33	33	33	33	33
34	34	34	34	34	34	34
35	35	35	35	35	35	35
36	36	36	36	36	36	36
37	37	37	37	37	37	37
38	38	38	38	38	38	38
39	39	39	39	39	39	39
40	40	40	40	40	40	40
41	41	41	41	41	41	41
42	42	42	42	42	42	42
43	43	43	43	43	43	43
44	44	44	44	44	44	44
45	45	45	45	45	45	45
46	46	46	46	46	46	46
47	47	47	47	47	47	47
48	48	48	48	48	48	48
49	49	49	49	49	49	49
50	50	50	50	50	50	50
51	51	51	51	51	51	51
52	52	52	52	52	52	52
53	53	53	53	53	53	53
54	54	54	54	54	54	54
55	55	55	55	55	55	55
56	56	56	56	56	56	56
57	57	57	57	57	57	57
58	58	58	58	58	58	58
59	59	59	59	59	59	59
60	60	60	60	60	60	60
61	61	61	61	61	61	61
62	62	62	62	62	62	62
63	63	63	63	63	63	63
64	64	64	64	64	64	64
65	65	65	65	65		

```
Query Match          98.2%;      Score 168;  DB 5;  Length 407;
Best Local Similarity 99.4%;      Pred. No. 2.02e-359;
Matches 168; Conservative 0; Mismatches 1;
```

[illegible]

Search completed: Tue Dec 22 05:24:26 1998
Job time : 292 secs.

CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
 CC comprising at least two functional regions - a binding domain with
 CC affinity for a ligand and a second effector peptide portion that is
 CC chemically or biologically active. They may further comprise a linker
 CC peptide between the 2 domains. The oligonucleotides are also designed so
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned
 CC in, or flanking, the unpredicted or variant residues. These residues
 CC confer some degree of conformational rigidity to the peptides. The TSARs
 CC or compps. comprising a TSAR binding domain can be used in vivo to
 CC deliver a chemically or biologically active moiety, eg. metal ion,
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
 CC cell. They can also replace the function of macromolecules, eg.
 CC monoclonal or polyclonal antibodies and therefore circumvent the need
 CC for complex methods of hybridoma formation or in vivo antibody
 CC production. The TSARs are easily characterised and have designed activity
 CC allowing direct and rapid detection in a screening process.
 SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 17.0%; Score 29; DB 12; Length 114;
 Best Local Similarity 3.7%; Pred. No. 1.74e-04;
 Matches 4; Conservative 30; Mismatches 75; Indels 0; Gaps 0;

Db 6 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 65
 QY 3 TCTGACGATCGCTGTCACCTTTGCGAAGTCTATTCCCTCCCTGCTGTCGCTT 62
 Db 66 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 114
 QY 63 TGTTCAGGTGGGATGATTTCATGTAGAGGTGAGAGAGGCTGCTAT 111

RESULT 8
 ID Q70468 standard; DNA; 114 BP.
 AC Q70468;
 DT 05-APR-1995 (first entry)
 DE Generic DNA sequence to generate a random TSAR peptide library.
 KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; generic; ss.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT misc_feature 55..60
 FT /*tag= a
 FT /note= "this sequence represents 'Z'; Z can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"
 PN W09418318-A.
 PD 18-AUG-1994.
 PF 01-FEB-1994; U00977.
 PR 01-FEB-1993; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.
 PA (UYNC-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Kay BK;
 DR WPI: 94-279739/34.
 DR P-PSDB: R65154.
 PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain
 PS Disclosure; Page 35; 255pp; English.
 CC Q70468 is a generic DNA sequence used to generate random TSAR (Totally
 CC Synthetic Affinity Reagents) peptides. This generic formula can also be
 CC represented as follows: X(NNB)11(TGC)(NNB)62(NNB)7(TGC)(NNB)10Y. X
 CC and Y are flanking restriction sites (X is not the same as Y) that are
 CC not specified further. Other generic sequences are shown in Q70466-68.
 CC Other specific peptides generated by these generic sequences are shown in
 CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
 CC comprising at least two functional regions - a binding domain with
 CC affinity for a ligand and a second effector peptide portion that is
 CC chemically or biologically active. They may further comprise a linker
 CC peptide between the 2 domains. The oligonucleotides are also designed so
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned
 CC in, or flanking, the unpredicted or variant residues. These residues

CC confer some degree of conformational rigidity to the peptides. The TSARs
 CC or compps. comprising a TSAR binding domain can be used in vivo to
 CC deliver a chemically or biologically active moiety, eg. metal ion,
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
 CC cell. They can also replace the function of macromolecules, eg.
 CC monoclonal or polyclonal antibodies and therefore circumvent the need
 CC for complex methods of hybridoma formation or in vivo antibody
 CC production. The TSARs are easily characterised and have designed activity
 CC allowing direct and rapid detection in a screening process.
 SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 17.0%; Score 29; DB 12; Length 114;
 Best Local Similarity 4.6%; Pred. No. 1.74e-04;
 Matches 5; Conservative 30; Mismatches 74; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 62
 QY 24 TTTGCGAAGTCTATTCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 83
 Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 111
 QY 84 CATGTAGAAGTGGGAAGAGGCTGCTATNACTCTTTGGATGGGAGTCTG 132

RESULT 9
 ID Q70469 standard; DNA; 114 BP.
 AC Q70469;
 DT 07-APR-1995 (first entry)
 DE Generic DNA sequence to generate a random TSAR peptide library.
 KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; generic; ss.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT misc_feature 55..60
 FT /*tag= a
 FT /note= "this sequence represents 'Z'; Z can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"
 PN W09418318-A.
 PD 18-AUG-1994.
 PF 01-FEB-1994; U00977.
 PR 01-FEB-1993; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.
 PA (UYNC-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Kay BK;
 DR WPI: 94-279739/34.
 DR Identifying proteins or peptide(s) which bind a ligand - by
 DR screening a recombinant vector library expressing fusion proteins
 DR comprising a binding domain and an effector domain
 PS Disclosure; Page 35; 255pp; English.
 CC Q70469 is a generic DNA sequence used to generate random TSAR peptide
 CC This generic formula can be represented as follows: X(TGC)(NNB)10-
 CC (TGC)(NNB)62(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking restriction
 CC sites (X is not the same as Y) that are not specified further. This
 CC sequence generates peptides that are cloverleaf in structure. Other
 CC generic sequences are shown in Q70465-68. Other specific peptides
 CC generated by these generic sequences are shown in R65150-54. TSARs are
 CC concatenated heterofunctional proteins or peptides, comprising at least
 CC two functional regions - a binding domain with affinity for a ligand and
 CC a second effector peptide portion that is chemically or biologically
 CC active. They may further comprise a linker peptide between the 2 domains.
 CC The oligonucleotides are also designed so that the expressed peptide
 CC contains 2 or 4 cysteine residues positioned in, or flanking, the
 CC unpredicted or variant residues. These residues confer some degree of
 CC conformational rigidity to the peptides. The TSARs or compps. comprising
 CC a TSAR binding domain can be used in vivo to deliver a chemically or
 CC biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
 CC or enzyme, to the specific target or on the cell. They can also replace
 CC the function of macromolecules, eg. monoclonal or polyclonal antibodies
 CC and therefore circumvent the need for complex methods of hybridoma
 CC formation or in vivo antibody production. The TSARs are easily

```

OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69 /*tag= a
FT /function-multiple cloning site
FT primer_bind 187..204 /*tag= b
PN EP-285123-A.
PD 05-MAY-1988. 105163.
PF 30-MAR-1988; US-034819.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure: p; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 18.7%; Score 32; DB 1; Length 204;
Best Local Similarity 6.6%; Pred. No. 4.12e-06;
Matches 6; Conservative 48; Mismatches 37; Indels 0; Gaps 0;

Db 95 rrmrbvnyrdvnrdaawccyrvskycynachhddhvybbvnyvnhnncnc 154
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 74 GGATGTAATTCATGTAGAGGTGGAAGAGCGTCTATNACTTTTGGATGGAGTCTGG 133
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 155 cbnhvchvnhbhnwavyvrharrddvh 185
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 134 CAAGAGGAATTTGAAGATAAAATAATAAT 164

RESULT 6
ID Q70465 standard; DNA; 114 BP.
AC Q70465;
DE 05-APR-1995 (first entry)
KW Generic DNA sequence to generate a random TSAR petide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60 /*tag= a
FT /note= "this sequence represents 'z'; z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
DR P-PSDB; R65150 and R65151.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure: Page 35; 255pp; English.
CC Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)12(TGC)(NNB)11Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC Q70466-68.

```

```

PS Disclosure: Page 35; 255pp; English.
CC Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)12(TGC)(NNB)11Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC Q70466-68.
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody
CC production. The TSARs are easily characterised and have designed
CC activity allowing direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 17.5%; Score 30; DB 12; Length 114;
Best Local Similarity 4.5%; Pred. No. 5.05e-05;
Matches 5; Conservative 31; Mismatches 76; Indels 0; Gaps 0;

Db 3 bnnbnbnbnbnbnbtgcnnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbn 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 3 TCTGAAGATGCGGTGTCAGTTGCGAAGTCTCTATTCCTCCTCTGTCGTGCTT 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 bnnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbtgcnnbnbnbnbn 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 63 TGTTGAGTGGGATGATTAATTCATGTAGAAGGTGGAAGGCTGCTATNAC 114

RESULT 7
ID Q70467 standard; DNA; 114 BP.
AC Q70467;
DE 05-APR-1995 (first entry)
KW Generic DNA sequence to generate a random TSAR petide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60 /*tag= a
FT /note= "this sequence represents 'z'; z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
DR P-PSDB; R65153.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure: Page 35; 255pp; English.
CC Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)12(TGC)(NNB)11Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC Q70466-68.

```

SQ	Sequence	172 BP;	46 A;	28 C;	48 G;	49 T;
	Query Match	100.0%;	Score 171;	DB 20;	Length 172;	
	Best Local Similarity	100.0%;	Pred. No. 1.28e-94;			
	Matches	172; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
D b	1	gatctgaacgatacgccgtgtccagtttgcgaagtccattccccctcctctctctgttgatgcc	60			
Q y	1	GATCTGAACGATCGCGTGTCCAGTTTCGGAAGTCTATTCCCCTCCTCTGTGGTGCC	60			
D b	61	ttgttcaggatggggatgtatttcattcatgaagggtgaaagaagcgctgactaactctttg	120			
Q y	61	TTTTGTCAGCTGGGGATGTATTTCATGTAGAAGGTGAAAGAGCGTCTA TNACTCTTTG	120			
D b	121	gatgggagctcgcgaagagaaaattgggaagataaaaaataaataaagtgaag	172			
Q y	121	GATGGGAGCTGC CAAGAGAAAATTGGAAGATAAAAAATAATAAAGTGAAA	172			

RESULT 2

ID Q51745 standard; cDNA; 91 BP.
AC Q51746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI: 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PT Claim 3; Page 14; 23pp; English.
PS Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 G; 4 T;

```

Query Match      23.43; Score 40; DB 9; Length 91;
Best Local Similarity 0.0%;
Pred. No. 1.27e-10;
Matches 0; Conservative 45; Mismatches 5; Indels 0; Caps 0;

Db 11 ssvshyvvvhhvshhshhvvhhvshvsvvvvhhvhhvhhvhhvhhvsvv 60
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 94 CTTTCATCATGAATACATCCCACCTCTGAACAAAGGCACACACAGGAGG 40
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT      3
ID          N81164 standard; DNA; 204 BP.
AC          N81164;
DT          08-NOV-1990 (first entry)
DE          Base substituted E.coli beta-galactosidase alpha-fragment.
KW          E.coli beta galactosidase alpha-fragment; base substitutions; ss.
KW          Escherichia coli.
FH          Key      Location/Qualifiers
FT          misc_feature      19..69
FT          /tag= a
FT          /function= multiple cloning site
FT          primer_bind      187..204
FT          /*tag= b
PN          EP-285123-A.
PD          05-MAY-1988.
PF          30-MAR-1988. 105163.
PR          03-APR-1987; US-034819.
PI          (SUSO) SUOMEN SOKERI OY.
PI          Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;

```

WPI:	88-279927/40.
DR	Introducing random point mutations into nucleic acids -
PT	by prepn of single stranded template, annealing a primer,
PT	misincorporation, completion of molecules and screening.
PS	Disclosure; p; English.
CC	Random point mutations were introduced into the alpha fragment of
CC	E.coli beta-galactosidase. The wild type sequence was obtained as a
CC	single stranded template and an oligonucleotide was hybridised to
CC	it to generate a popn of DNA molecules which terminate at all
CC	possible nucleotide positions within a specified region. The
CC	variable 3' ends generated in this way are used as primers for
CC	reverse transcriptase. Nucleotides are misincorporated by the
CC	transcriptase and the molecules are completed to forms that can be
CC	amplified and then expressed in a suitable host-vector system.
CC	The sequence covers all 176 diff base substitutions, most of which
CC	occurred singularly in any given mutant.
CC	See also P80575.
SQ	Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

```
Query Match      21.1%; Score 36; DB 1; Length 204;
Best Local Similarity 9.8%; pred. No. 2.44e-08;
Matches 11; Conservative 58; Mismatches 42; Indels 1; Gaps 1;
```

D b	74	aaycdchvgcgymrtthhyrrmbvnyrdynrdaawwccyr-rsvkydcycnakh	132
		: : :	: : : : : : : : : : : : :
C p	152	ATCTTCCAATTCTCTGTCCAGACTCCCATTCAAAGAGINATGACGCCTTCTTCCACC	93
D b	133	ddhyvybbbyvnhnancncbhnvhcnvbnhnrwayvrhdaradv	184
		: : : : : :	: : : : : : : : : : : : : : :
C p	92	TTCATCATGAATAATCATCTCCCACTGTAAACAAGSCACCACTCAGGAGGGG	41

RESULT 4

ID Q51746 standard; cDNA; 91 BP.

AC Q51746;

DT 31-MAY-1994 (first entry)

DE Oligonucleotide probe MK14-A

KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;

KW ss.

OS Synthetic.

PN EP-571911-A.

PD 01-DEC-1993.

PF 24-MAY-1993; 108325.

PR 26-MAY-1992; US-889651.

PA (BECT) BECTON DICKINSON CO.

PI Shank DD, Spears PA;

DR WPI: 93-378844/48.

PT New oligo:nucleotide probes specific for Mycobacteria - used for

PT detection and amplification of Mycobacteria nucleic acid in

PT samples

PS Claim 3; Page 14; 23pp; English.

CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14

CC (Q51735). It hybridized to all spp. of mycobacteria tested, but

CC cross reacted to a few non-mycobacterial spp. the probe may

CC be useful as an initial screen for mycobacterial infection.

CC See also Q51735-45 and Q51747-59.

CC Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

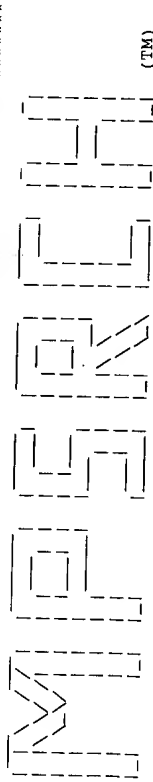
```

Query Match      18.7%   Score 32;   DB 9;   Length 91;
Best Local Similarity 0.0%   Pred. No. 4.12e-06;
Matches 0;   Conservative 8;   Mismatches 0;   Gaps 0;

Db      13   vhsyvvvhhshshshhvvhhvhsvvvhhvvvhhvvhhvhyvsv 60
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      13   GCCGTGTCAGTTTGCAAGTCCTATATCCCTCCTCCTGTCGTCGTCGC 60

```

RESULT	5
ID	N81164 standard; DNA; 204 Bp.
AC	N81164;
DT	08-NOV-1990 (first entry)
DE	Base substituted E.coli beta-galactosidase alpha-fragment.
KW	E.coli beta galactosidase alpha-fragment; base substitutions; ss.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 05:24:45 1998; MasPar time 42.32 Seconds
Tabular output not generated. 553.013 Million cell updates/sec

Title: >US-08-530-112A-2965
Description: (1-172) from US08530112A.seq
Perfect Score: 171
N.A. Sequence: 1 GATCTGAACGATGCCGTGTC.....AAATAATATATAGTGAAA 172
Comp: CTAGACTTGTCTACGCCACAG.....TTTATTATTATTCACCTTT

Scoring table: TABLE default

Gap 5

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: n-genes32

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 7.450; Variance 4.668; scale 1.596

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match %	Length	DB ID	Description	Pred. No.
1	171	100.0	172	20	T21965	
2	40	23.4	91	9	Q51746	Human gene signature 1.28e-94
3	36	21.1	204	1	N81164	Oligonucleotide probe 1.27e-10
4	32	18.7	91	9	Q51746	Base substituted E.co 2.44e-08
5	32	18.7	204	1	N81164	Oligonucleotide probe 4.12e-06
6	30	17.5	114	12	Q70465	Base substituted E.co 4.12e-06
7	29	17.0	114	12	Q70465	Generic DNA sequence 5.05e-05
8	29	17.0	114	12	Q70467	Generic DNA sequence 1.74e-04
9	28	16.4	114	12	Q70468	Generic DNA sequence 1.74e-04
10	27	15.8	114	12	Q70469	Generic DNA sequence 5.94e-04
11	26	15.2	178	32	T76405	Human endothelin-1 an 1.98e-03
12	26	15.2	114	12	Q70466	Generic DNA sequence 6.57e-03
13	24	14.0	114	12	Q70470	Generic DNA sequence 6.88e-02

SUMMARIES

ALIGNMENTS

RESULT 1

ID T21965 standard; cDNA to mRNA; 172 BP.
AC T21965;
DT 05-AUG-1996 (first entry)
DE Human gene signature HUMGS03508
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PN 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATSU) MATSUBARA K.
PA (OKUB) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues.
PS Claim 1; Page 1012; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.

Human interleukin 8 a 6.88e-02
TSAR-9 library genera 2.17e-01
TSAR-9 library genera 2.17e-01
DC43 TSAR library gen 2.17e-01
DC43 TSAR library gen 2.17e-01
DC43 TSAR library gen 6.71e-01
DC43 TSAR library gen 6.71e-01
Generic DNA sequence 6.71e-01
Human IL-8 receptor-a 2.03e+00
Human eosinophil deri 2.03e+00
TSAR-9 library genera 2.03e+00
TSAR-9 library genera 2.03e+00
Generic DNA sequence 2.03e+00
Generic DNA sequence 2.03e+00
Generic DNA sequence 6.03e+00
Generic DNA sequence 6.03e+00
Human vascular cell a 6.03e+00
Chymase antisense oli 6.03e+00
Substance P antisense 6.03e+00
CC288_9 cDNA clone. 6.03e+00
Canola palmitoyl-ACP- 6.03e+00
Probe for Candida alb 1.74e-01
Generic DNA sequence 1.74e-01
Generic DNA sequence 1.74e-01
Human gene signature 1.74e-01
Pyrococcus furiosus R 1.74e-01
Fungus-responsive pot 1.74e-01
DmORF1 potassium chan 1.74e-01
Human lysosomal membr 1.74e-01
Sequence encoding tum 1.74e-01
Human neuropeptide Y- 1.74e-01

NID KEYWORDS 93337381
SOURCE HTG.
ORGANISM human.

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS Primates; Catarrhini; Hominiidae; Homo.
TITLE 1 (bases 1 to 216021)
JOURNAL Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
Mason,T.M., Brandon,R., Kim,U.T., Kerlavage,A.R. and Venter,J.C.
Homo sapiens Chromosome 16 BAC clone C1F987SK-A-952F10
Unpublished

REFERENCE 2 (bases 1 to 216021)
AUTHORS Adams,M.D. and Loftus,B.J.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, Email:
[bjloftus@tigr.org](#)

REFERENCE 3 (bases 1 to 216021)
AUTHORS Adams,M.D. and Loftus,B.J.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA

COMMENT On Jul 24, 1998 this sequence version replaced gi:3241936.
Address all correspondence to: Mark Adams The Institute for Genomics,
Research 9712 Medical Center Dr., Rockville, MD 20850, USA e-mail
address: [humgen@tigr.org](#). The orientation of the sequence is from
SP6 end to T7 end. Genes were identified by a combination of five
methods including: XGRAIL ([available](#)) by anonymous ftp from
[arthur.epm.ornl.gov](#), GeneFinder (Phil Green, University of
Washington), GENSCAN (Chris Burge,
[http://genome.stanford.edu/~chris/GENSCANW.html](#)) searches of the
complete sequence against a peptide database, and the Human gene
Index database at TIGR ([http://www.tigr.org/tldb/bg1/bgi.html](#)).
Genes without predicted homology having spliced EST hits are termed
'unknown gene product'. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, [http://genome.wustl.edu/eddy/tRNAscan.SE/](#)).
Location/Qualifiers
1..216021
 '/organism="Homo sapiens"
 '/db_xref="taxon:9606"
 '/chromosome="16"
 '/map="A:16q21-22"
 '/clone="A-952F10"
 ' 27865..27872
 /note="T766, STS1-CSRL-27g3-uA/CSRL-27g3-uZ, Chr. -, Homo
 sapiens"
 /db_xref="dbSTS:C02280"
 73826..73943
 /note="T7608, STS1-CSRL-24g1-uA/CSRL-24g1-uZ, Chr. -, Homo
 sapiens"
 '/db_xref="dbSTS:C02122"
 175801..175945
 /note="I6084, CHLC.GCT10B02, Chr. -, Homo sapiens"
 '/db_xref="dbSTS:G09703"
 175810..175945
 /note="I6316, CHLC.GCT15C04, Chr. -, Homo sapiens"
 '/db_xref="dbSTS:C09935"
 199463..199572
 /note="9824, WI-3555, Chr. 16, Homo sapiens"
 '/db_xref="dbSTS:G04338"

BASE COUNT 60960 A 51778 C 49172 G 53987 T 124 others
Origin

Query Match 21.6% Score 37; DB 26; Length 216021;
Best Local Similarity 16.8%; Pred. No. 2,59e+08;

Matches 21; Conservative 62; Mismatches 41; Indels 1; Gaps 1;

Dbb 1436 WYGSYYCCYCGRARRRGKKKYMMSMRARRRSGAGKKKYCYCYYYYYYCM 1495
::|.:||: : : : :::::::::::::: | : ::::: : : ::::|
CP 148 TCCAATTTCGTGCAGACTCCCATCCAAGAAGTNATGACGCCCTTTCCACCCTTC 89
||||: : : ||: : : : : ||: : : : : ||: : : : : ||:

Dbb 1496 GRAMMAAMYKRRCSCAAMYTMGARAKTYIIYAAR-PCCARSYKKMYAMAANTWC 1554
| |||: : : ||: : : : : ||: : : : : ||: : : : :

CP	88	ACATGAATACATCCCACTGACCAAGCACACACAGAGAGGAGGATAGACTTC	29
Db	1555	AAAAA	1559
CP	28	GCAAA	24
RESULT	3		
LOCUS	AF012089	10772 bp	DNA INV 05-AUG-1997
DEFINITION	Drosophila melanogaster cysteine proteinase-1 (Cp1) gene, complete cds, and phenylalanyl tRNA synthetase gene, partial cds.		
ACCESSION	AF012089		
NID	92305220		
KEYWORDS	fruit fly, Drosophila melanogaster		
ORGANISM	Eukaryote; mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 4546 to 4553) Gray, Y.H., Tanaka, M.M. and Sved, J.A.		
AUTHORS	P-element-induced recombination in Drosophila melanogaster: hybrid element insertion		
TITLE	Genetics 144 (4), 1601-1610 (1996)		
JOURNAL	MEDLINE 97132596		
REFERENCE	2 (bases 1 to 10772) Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.		
AUTHORS	Structure of the cysteine proteinase (Cp1) gene of Drosophila melanogaster and associated mutational effects		
TITLE	Unpublished		
JOURNAL	3 (bases 1 to 10772) Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.		
REFERENCE	Submitted (30-JUN-1997) School of Biological Sciences, University of Sydney, Biology A12, Sydney University, NSW 2006, Australia		
AUTHORS	Location/Qualifiers		
TITLE	1..10772		
FEATURES	/organism="Drosophila melanogaster"		
source	/db_xref="taxon:7227"		
mrna	join(872..1000,2310..2426,6476..6690,6751..7707)		
gene	/gene="Cp1"		
exon	/product="cysteine protease"		
intron	872..7707		
exon	/gene="Cp1"		
intron	872..1000		
exon	/gene="Cp1"		
intron	/number=1		
exon	1001..2309		
intron	/gene="Cp1"		
exon	2310..2426		
intron	/gene="Cp1"		
exon	/number=2		
intron	join(2328..2426,6476..6690,6751..7462)		
exon	/gene="Cp1"		
intron	/db_xref="P10:92305221"		
exon	/product="cysteine proteinase-1"		
intron	/codon_start=1		
exon	/db_xref="P10:92305221"		
intron	/translation="MRTAVLLPLALLAVAQVSFADYVMEEMHFKLEHRRNYODET		
exon	EERRRLKTEFNKHKIAKNORPAEGKVSFKLVAKRYADLLHHEFRLMNGFNTLLHKK		
intron	QLRAADSEFGVTFISPAHYTLPEKSYDMRTKGAVTAAVADOGCSNAFSTGALDECO		
exon	HFRSGYVLSLEONLVDCSTKGGNGGIMONAFRYIKDNGSIDREKSPYEALD		
intron	DSCFHNKGTAVDRGFTDIPQODEKKMAENATYGPVSALDSHESFOFSGVYN		
exon	EPOCDANLDHGVLYVVGFTDSEGEDYMLVKNMGFTWGDGFKLKNKNOGCIAS		
intron	ASSPLV"		
exon	2427..6475		
intron	/gene="Cp1"		
exon	4546..4553		
intron	/gene="Cp1"		
exon	/note="insertion site of P[casPer](50C)"		
intron	/citation=[1]		
exon	6476..6690		
intron	/gene="Cp1"		
exon	/number=3		


```

Intron
    6691..6750
    /gene="Cpl"
    6751..7707
    /gene="Cpl"
    /number=4
    join(<8110..9300,9370..>9532)
    /product="phenylalanyl tRNA synthetase"
    join(8110..9300,9370..>9532)
    /note="potential orf"
    /codon_start=1
    /product="phenylalanyl tRNA synthetase"
    /db_xref="PID:Q2305222"
    /translation="MULTIRVCGARHMLKSTRCLASAPAKSPSPPOLEVGSTYA
    TDCMTVTPKILSYVGANKRLQDHPILSTRORIVNFTGAYRNQGNPLFSYDDQN
    PVTVOQNDNLIPADHVSROSDCTYINQOHLRAHTAHQVELISGLDNFLVYG
    EYVRDEIDSTHLPVHQADAVRLVTKDLFERNPGLFEETWSTGLADPKLLPH
    SSWTKNSPATRAKVKLMEHEMKHVLGLKDLFGPKIKRWVDYTPFPQSPWELEI
    YFQDNMLEVGGCIMERHEILQSRGSHVIGTARGVGLERLAVLFDIPDIRLMSNDS
    GFLSOSSEKDHLNPKYKPISSHVPOCNDISFPLPDIEVDAGFSPNDFYDLRSVAG
    DMEQISLVDRKFKHPTKGSVCFRIVYRHEMERLTDAEVNEIHKQIASASVDSFNVQ
    IX

BASE COUNT      2929 a      2357 c      2282 g      3046 t      158 others
ORIGIN

Query Match      17.5%: Score 30; DB 18; Length 10772;
Best Local Similarity 17.9%: Pred. No. 4,84e-04;
Matches 21; Conservative 57; Mismatches 37; Indels 2; Gaps 2;

Db 1704 WKMMWTTWKMMAMKTYRTMMWMMKMYTSTTTTSAMMMWYMTWTYKMMAYAMMMKMMWT 1763
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 57 TGCCTTGTTCAGTGGGAGTGTATTCATGTAGAGGTGGAAGAGCTGC-TATNACT 115
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1764 RTWARMAMASWARKMKWTSAAAYISANRKMKW-AYRAMKRTMMAMAMKMKRKAAM 1819
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 116 CTTTGATGGAGTGTGCAAGGAATGGAAGATTAATAATATAAGTGAA 172
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 4
LOCUS 128278 215 bp DNA PAT 30-OCT-1996
DEFINITION Sequence 5 from patent US 5569830.
ACCESSION 128278
NID 91819054
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 215)
AUTHORS Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
TITLE Plant inhibitors of fungal polygalacturonases and their use to
JOURNAL Patent: US 5569830-A 5 29-OCT-1996;
FEATURES
    source
        1..215
        /organism="unknown"
BASE COUNT      15 a      8 c      25 g      26 t      141 others
ORIGIN

Query Match      17.0%: Score 29; DB 21; Length 215;
Best Local Similarity 18.8%: Pred. No. 1,85e-03;
Matches 30; Conservative 55; Mismatches 74; Indels 1; Gaps 1;

Db 26 GNTTSSTTTDCNRTWCVDCTDITTYRVNNDSGHKKYSSANYNGGNNYGAARKHYHTTN 85
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 10 GATGCCCTGTCAGTTCGCAAGTCTATTCCTCCCTCGTGTGCTTGTTCAG 69
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 86 VSGADSTVTDSDYN-ASGTSSNGTGDGNSGADSYGSSKTAAMSRTNKTANNVDSR 144
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 70 GTGGGGATGTATTTCATGTAGAGGTGGAAGAGCTGTATNACTTTTGATGGAGGT 129
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 145 NMGDASVSDKNTKKHAKNSADGKVGSKNNGDRNNRYGTG 184
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 130 CTGGCAAGAGGAATGTGAAGATTAATAATATAAGTG 169
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

RESULT 5
LOCUS 166494 7218 bp DNA PAT 23-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
NID 92724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 7218)
AUTHORS Dorner,F., Schefflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
    source
        1..7218
        /organism="unknown"
BASE COUNT      1944 a      1491 c      1486 g      1929 t      368 others
ORIGIN

Query Match      16.4%: Score 28; DB 21; Length 7218;
Best Local Similarity 17.4%: Pred. No. 6,95e-03;
Matches 8; Conservative 29; Mismatches 9; Indels 0; Gaps 0;

Db 1057 AGCTTGCGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1102
    |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 22 AGTTGCGAAGTCTCATTCCTCCCTCCTCGTGTGCTGCTTTGTC 67
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 6
LOCUS AF012089 10772 bp DNA INV 05-AUG-1997
DEFINITION Drosophila melanogaster cysteine proteinase-1 (Cpl) gene, complete
ACCESSION AF012089
NID 92305220
KEYWORDS
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
REFERENCE
1 (bases 4546 to 4553)
AUTHORS Gray,Y.H., Tanaka,M.M. and Sved,J.A.
TITLE P-element-induced recombination in Drosophila melanogaster: hybrid
JOURNAL Genetics 144 (4), 1601-1610 (1996)
MEDLINE 97132596
REFERENCE
2 (bases 1 to 10772)
AUTHORS Gray,Y.H.M., Sved,J.A., Preston,C.R. and Engels,W.R.
TITLE Structure of the cysteine proteinase (Cpl) gene of Drosophila
JOURNAL melanogaster and associated mutational effects
REFERENCE
3 (bases 1 to 10772)
AUTHORS Gray,Y.H.M., Sved,J.A., Preston,C.R. and Engels,W.R.
TITLE Direct Submission
JOURNAL Submitted (30-JUN-1997) School of Biological Sciences, University
of Sydney, Biology A12, Sydney University, NSW 2006, Australia
FEATURES
    source
        1..10772
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        join(872..1000,2310..2426,6476..6690,6751..7707)
        /gene="Cpl"
        /product="cysteine protease"
        872..7707
        /gene="Cpl"
        872..1000
        /gene="Cpl"
        /number=1
        1001..2309
        /gene="Cpl"
        2310..2426
        /gene="Cpl"

```



```

MEDLINE      84010860
FEATURES
Source        Location/Qualifiers
1..2221
misc_feature  /organism="Capra hircus"
              /db_xref="taxon:9925"
              311..318
CAAT_signal   /note="conserved sequence in embryonic globine genes"
              340..344
              /note="pot. CAAT-box"
              394..400
TATA_signal   /note="pot. TATA-box"
              425
misc_RNA      /note="pot. CAP-site"
              425..569
              /number=1
RBS           441..447
              /note="pot. Shine-Dalgarno sequence"
              join(478..569,686..908,1948..2076)
CDS           /codon_start=1
              /product="epsilon I globin"
              /db_xref="pid:9970"
              /db_xref="SWISS-PROT:P02102"
              /translation="MVHFTAEKKAITGLMKVNEEAGGALGRLLVYPWTOREFD
              SFGNLSSASALMGNPKVAHGKRVLTSGEALKNLDNKGAFKSELHCDKLHVDP
              E
              NFRLLGNVIVIIATHTGREFTPDQAAMQKLVSVATALAHKYH"
              570..685
              /number=1
              /note="e-I mRNA"
              join(478..569,686..908,1948..2076)
              /note="beta-globin epsilon-I"
              /codon_start=1
              /db_xref="PID:g164144"
              /translation="MVHFTAEKKAITGLMKVNEEAGGALGRLLVYPWTOREFD
              SFGNLSSASALMGNPKVAHGKRVLTSGEALKNLDNKGAFKSELHCDKLHVDP
              E
              NFRLLGNVIVIIATHTGREFTPDQAAMQKLVSVATALAHKYH"
              <478..569
              /note="beta-globin epsilon-I"
              /number=1
              /note="e-I intron a"
              686..908
              /number=2
              909..1947
              /note="e-I intron II"
              /number=2
              1095..1107
              /note="13 bp imperfect direct repeat 1"
              1102..1108
              /note="7 bp imperfect direct repeat 2"
              1109..1121
              /note="13 bp inverted repeat 1"
              1109..1328
              /organism="Capra hircus"
              /insertion_seq=""
              /note="220 bp insertion sequence"
              1124..1236
              /note="highly repetitive ungulate sequence"
              1317..1329
              /note="13 bp inverted repeat 1"
              1322..1334
              /note="13 bp imperfect direct repeat 1"
              1329..1335
              /note="7 bp imperfect direct repeat 2"
              1948..>2076
              /note="pot. exon 3"
              /number=3
              2168..2173
              /note="pot. polyA signal"
              2168..2173
              /number=3
              462 c 526 g 603 t 1 others
BASE COUNT 629 a 462 c 526 g 603 t 1 others
ORIGIN
Query Match 13.5%: Score 23; DB 19; Length 2221;
Best Local Similarity 69.0%: Pred. No. 3,71e+00;
Matches 49; Conservative 0; Mismatches 21; Indels 1; Gaps 1;
polyA_signal
Db 1685 TAGAAGGAGGAAAGGCACTAAATAATGTAATAGAGAGAGCAAG-GGATATAGG 1743
      ||||| ||| ||||| ||| ||| ||||| ||||| ||| |||
QY 88 TAGAAGGTGGAAGAGGCTGCTATNACTCTTGATGGGAGTCTGGCAAGAGAAATTGG 147
      ||| ||||| ||| ||| ||||| ||||| ||| |||
Db 1744 CAGACAAATA 1754
      ||| ||||| |||
QY 148 AAGATAAATA 158

```

```

RESULT 13
LOCUS   GOTHBEI 2221 bp DNA MAM 10-DEC-1984

```

```

DEFINITION  Goat embryonic beta-globin epsilon-I, complete gene and flanks.
ACCESSION   K01713
NID         g164143
KEYWORDS    beta-globin; epsilon-I beta-globin; globin; hemoglobin.
SOURCE      Goat genomic DNA, Charon 4A library, clone pI-VX-epsilon-5'.
ORGANISM    Capra hircus
            Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Caprinae; Capra.
REFERENCE   1 (bases 1 to 2221)
AUTHORS     Shapiro,S.G., Schon,E.A., Townes,T.M. and Lingrel,J.B.
TITLE       Sequence and linkage of the goat epsilonI and epsilonII beta-globin
            genes
JOURNAL     J. Mol. Biol. 169, 31-52 (1983)
MEDLINE     84010860
COMMENT     [1] reports the linkage arrangement of the beta-globin genes in the
            goat which is 5'-epsilon-I - epsilon-II - psi-beta-X - beta-C-3'.
            [1] gives the sequences of the goat epsilon-I and epsilon-II
            beta-globin genes, which were compared to the globin genes of other
            organisms and other goat globin genes. Some degree of homology was
            found among found among all of them, suggesting a common ancestry.
FEATURES
Source      Location/Qualifiers
1..2221
            /organism="Capra hircus"
            /db_xref="taxon:9925"
            /note="e-I mRNA"
            join(478..569,686..908,1948..2076)
            /note="beta-globin epsilon-I"
            /codon_start=1
            /db_xref="PID:g164144"
            /translation="MVHFTAEKKAITGLMKVNEEAGGALGRLLVYPWTOREFD
            SFGNLSSASALMGNPKVAHGKRVLTSGEALKNLDNKGAFKSELHCDKLHVDP
            E
            NFRLLGNVIVIIATHTGREFTPDQAAMQKLVSVATALAHKYH"
            <478..569
            /note="beta-globin epsilon-I"
            /number=1
            /note="e-I intron a"
            686..908
            /number=2
            909..1947
            /note="e-I intron b"
            1948..>2076
            /note="beta-globin epsilon-I"
            /number=3
            462 c 526 g 603 t 1 others
BASE COUNT 629 a 462 c 526 g 603 t 1 others
ORIGIN
Query Match 13.5%: Score 23; DB 19; Length 2221;
Best Local Similarity 69.0%: Pred. No. 3,71e+00;
Matches 49; Conservative 0; Mismatches 21; Indels 1; Gaps 1;

```

```

Db 1685 TAGAAGGAGGAAAGGCACTAAATAATGTAATAGAGAGAGCAAG-GGATATAGG 1743
      ||||| ||| ||||| ||| ||| ||||| ||||| ||| |||
QY 88 TAGAAGGTGGAAGAGGCTGCTATNACTCTTGATGGGAGTCTGGCAAGAGAAATTGG 147
      ||| ||||| ||| ||| ||||| ||||| ||| |||
Db 1744 CAGACAAATA 1754
      ||| ||||| |||
QY 148 AAGATAAATA 158

```

```

RESULT 14
LOCUS     AC005369 74371 bp DNA PRI 01-AUG-1998
DEFINITION Homo sapiens chromosome 5, BAC clone 119j3 (LBNU H175), complete
            sequence.
ACCESSION AC005369
NID       g3367505
KEYWORDS  HTG.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.

```

```
REFERENCE 1 (bases 1 to 74371)
AUTHORS Kimerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE Sequencing of human chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 74371)
AUTHORS Rhee,D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 74371)
AUTHORS Kimerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
COMMENT Sequence submitted by:
DOE Joint Genome Institute.
FEATURES
source 1. 74371
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="5q"
/clone="11913"
/chromosome="5"
/note="LBNL H175"
893..1030
/repeat_region /rpt_family="Alu"
2285..2438
/rpt_family="Alu"
2818..2859
/note="(GT)21"
/rpt_type=tandem
/rpt_unit=GT
join(3246..3410,3721..3828)
/misc_feature
/standard_name="RLF"
/note="65% & 69% protein identity GenPept:U2377"
3431..3724
/rpt_family="Alu"
3707..3728
/note="(A)22"
/rpt_type=tandem
/rpt_unit=A
4366..4661
/rpt_family="Alu"
5327..5602
/rpt_family="Alu"
6386..6956
/rpt_family="L1"
6647..6684
/note="(CA)19"
/rpt_type=tandem
/rpt_unit=CA
7113..7373
/rpt_family="Alu"
complement(7830..8185)
/standard_name="possible repeat"
8258..8503
/rpt_family="Alu"
9070..9387
/rpt_family="Alu"
complement(9740..9845)
/rpt_family="MER42"
complement(10440..11015)
/rpt_family="Alu"
11950..12250
/rpt_family="Alu"
12057..12085
/note="(A)29"
/rpt_type=tandem

repeat_region /rpt_unit=A 12365..12645
repeat_region /rpt_family="Alu" 13727..13750
/note="(AC)12"
/rpt_type=tandem
/rpt_unit=AC 13783..14024
repeat_region /rpt_family="L1" 14175..14470
repeat_region /rpt_family="Alu" complement(14906..15259)
repeat_region /standard_name="possible repeat" 15300..15613
/rpt_family="Alu" 16671..16690
/note="(T)20"
/rpt_type=tandem
/rpt_unit=T
repeat_region complement(16675..16977)
/rpt_family="Alu" complement(16993..17085)
/rpt_family="MER42" complement(17678..18276)
/rpt_family="Alu" 19305..19583
/rpt_family="Alu" 19914..19945
/note="(GTT)8"
/rpt_type=tandem
/rpt_unit=GT
repeat_region complement(19943..20222)
/rpt_family="Alu" 21202..21496
/rpt_family="Alu" 21736..22035
/rpt_family="Alu" 22017..22038
/note="(A)22"
/rpt_type=tandem
/rpt_unit=A 22321..22457
/rpt_family="Alu" complement(22638..22981)
/standard_name="possible repeat" 23022..23326
/rpt_family="Alu" 23473..23761
/rpt_family="Alu" 23744..23767
/note="(A)24"
/rpt_type=tandem
/rpt_unit=A complement(24639..24694)
/rpt_family="MER42" complement(25349..25713)
/rpt_family="Alu" complement(25727..26471)
/rpt_family="Alu" 27191..27477
/rpt_family="Alu" 27774..28057
/rpt_family="Alu" 28040..28066
/note="(A)27"
/rpt_type=tandem
/rpt_unit=A 28712..28930
/rpt_family="MER20" complement(28769..28838)
/note="GRAIL 2 excellent exon, frame 0" 28987..29214
/rpt_family="Alu"
```

```

misc_feature      join(29485..29559,29598..29779)
                  /note="95% identity dbSTS:G14522 (SHGC-11312)"
misc_feature      29495..29976
                  /standard_name="A1027942"
                  /note="100% identity EST ov84a10.x1"
misc_feature      /note="GRAIL 2 excellent exon, frame 2"
repeat_region     /rpt_family="MIR"
                  /rpt_family="MIR"
misc_feature      complement(31573..31724)
                  /note="GRAIL 2 excellent exon, frame 1"
misc_feature      complement(32159..32232)
                  /note="GRAIL 2 excellent exon, frame 2"
repeat_region     32388..32488
                  /rpt_family="MLT1"
repeat_region     32617..32908
                  /rpt_family="Alu"
repeat_region     32977..33088
                  /rpt_family="MLT1"
repeat_region     complement(33670..33785)
                  /rpt_family="Alu"
misc_feature      complement(34021..34144)
                  /note="GRAIL 2 excellent exon, frame 2"
misc_feature      complement(35238..35331)
                  /note="GRAIL 2 excellent exon, frame 1"
repeat_region     36392..36663
                  /rpt_family="Alu"
misc_feature      36901..37222
                  /note="100% identity EST ou55c09.x1"
STS               /db_xref="dbEST:A1025011"
                  36901..37164
                  /standard_name="STSG-9983"
                  /db_xref="dbSTS:G26554"
misc_feature      complement(37595..37654)
                  /note="GRAIL 2 excellent exon, frame 0"
misc_feature      complement(join(38063..38218,38462..38578,38741..38995,
                  39071..39205,39532..39630,39935..40048,40300..40410,
                  40503..40661,41868..41972,42103..42225,42492..42569,
                  44379..44507))
                  /standard_name="histidyl-tRNA synthetase"
                  /note="78% 100% protein identity GenPept:018937"
misc_feature      complement(38069..38215)
                  /note="GRAIL 2 excellent exon, frame 0"
misc_feature      complement(38462..38578)

Note: remainder of annotations omitted.

Query Match      13.5%; Score 23; DB 26; Length 74371;
Best Local Similarity 20.7%; Pred. No. 3.71e+00;
Matches 6; Conservative 20; Mismatches 3; Indels 0; Gaps 0;

DB 15839 MRRAAWTKSMRRMMWMTWMMWRRRAA 15867
OY 138 AGGAATGTGAGATTAATAATAATAA 166

```

```

MEDLINE          93341463
REFERENCE        2 (bases 1 to 2766)
AUTHORS          Noelling,J.
TITLE            Direct Submission
JOURNAL          Submitted (02-NOV-1992) J. Noelling, Wageningen Agricultural
                  University, Bacterial Genetics Group, Dept of Microbiology,
                  Hesselink van Suchtelenweg 4, 6703 CT Wageningen, THE NETHERLANDS

FEATURES
Source           location/Qualifiers
                1..2766
                /organism="Methanobacterium thermoautotrophicum"
                /db_xref="taxon:2166"
                902..2402
                /note="FR-1a"
                937..1176
                /note="Orf1"
                /codon_start=1
                /transl_table=1
                /db_xref="PID:g44585"
                /translation="MGVKEDIRGOIIGALGADPEIINSPPELMAALPNGPDTCKSGD
                VELKASDAGVITADDPFRKSAEEVADTIYKAGL"
                complement(1516..2268)
                /note="Orf4"
                /codon_start=1
                /transl_table=1
                /db_xref="PID:g44586"
                /translation="MKNLKNQNDPEFYDARKLEATHODIKRIMESNOEYLDMLSN
                LKQILNSVYIISDDIENDLEQNMVNPCKMRECKEKFNGFIQNNSKLIKQHVSRK
                IIEERRELDIEIRKSAFPCDCICFKEVDSLFPKQNLIGSLQIDYNNKEDKPEISSI
                PEELMVSVLEPIISKORQOILKSMASEMTFTFALSELTLRGGNLLEFHIOKLLLESPL
                ILORHERGDYMIETKGYNLILMSNEFKYLYK"

BASE COUNT      781 a 548 c 583 g 854 t

ORIGIN
Query Match      12.9%; Score 22; DB 15; Length 2766;
Best Local Similarity 71.2%; Pred. No. 1.20e+01;
Matches 37; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

DB 719 TCCTGTGCAGCTCGATGTCAGATGGGAGTATTGAGGTTCACATGGA 770
OY 47 TCCTGTGTGCTGCTTGTTCAGGTGGGAGTATTTCATGTAAGGTGGA 98

```

Search completed: Tue Dec 22 05:19:16 1998
Job time : 310 secs.

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 438)
AUTHORS
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1550 Std Error: 0.00
Seq primer: -40m13 fwd. Et from Amersham
High quality sequence stop: 371.
Location/Qualifiers

FEATURES

source

1..438
/organism="Homo sapiens"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_image="162163"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
BASE COUNT 132 a 107 c 79 g 120 t
ORIGIN

Query Match 100.0%; Score 87; DB 17; Length 438;
Best Local Similarity 98.9%; Pred. No. 3.21e-74;
Matches 87; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 TTTAATGTAGAGATGATTTTATTAGACAATTCACATTTTACTATATTTTCCTT 60
|||||
Cp 88 TTTAGATGTAGAGATGATTTTATTAGACAATTCACATTTTACTATATTTTCCTT 29
|||||
Db 61 CTTAAACCTCACCATCAAGATTGATC 88
|||||
Cp 28 CTTAAACCTCACCATCAAGATTGATC 1
|||||

RESULT 5
LOCUS AA496035 455 bp mRNA EST 11-AUG-1997
DEFINITION zv72d06.s1 Soares total fetus Nb2HF8 9w Homo sapiens CDNA clone
759179 3', mRNA sequence.
ACCESSION AA496035
NID 92229356
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea;
Homo.

REFERENCE
1 (bases 1 to 455)
AUTHORS
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,
Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theisling, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE
WashU-Merck EST Project 1997
JOURNAL
Unpublished (1997)
COMMENT

Contact: Wilson RK
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. Et from Amersham
High quality sequence stop: 420.
Location/Qualifiers

FEATURES

source

1..455
/organism="Homo sapiens"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand CDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
GTTTACCATCTGAGTCGAGCGCGCTTAATTTTCTTTTCTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_lib="Soares total fetus Nb2HF8 9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
complement(<1..>455)
BASE COUNT 134 a 109 c 83 g 129 t
ORIGIN

Query Match 100.0%; Score 87; DB 8; Length 455;
Best Local Similarity 98.9%; Pred. No. 3.21e-74;
Matches 87; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 TTTAGATGTAGAGATGATTTTATTAGACAATTCACATTTTACTATATTTTCCTT 60
|||||
Cp 88 TTTAGATGTAGAGATGATTTTATTAGACAATTCACATTTTACTATATTTTCCTT 29
|||||
Db 61 CTTAAACCTCACCATCAAGATTGATC 88
|||||
Cp 28 CTTAAACCTCACCATCAAGATTGATC 1
|||||

RESULT 6
LOCUS AA032240 501 bp mRNA EST 10-MAY-1997
DEFINITION zK19a02.s1 Soares pregnant uterus NbHPU Homo sapiens CDNA clone
470954 3', mRNA sequence.
ACCESSION AA032240
NID 91502202
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea;
Homo.

REFERENCE
1 (bases 1 to 501)
AUTHORS
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisan, E.,
Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
TITLE
WashU-Merck EST Project
JOURNAL
Unpublished (1995)
COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1148 Std Error: 0.00

FEATURES

Seq primer: -40M13 fwd. from Amerisham
High quality sequence stop: 428.
Location/Qualifiers
1. 501

source

/organism="Homo sapiens"
/note="Organ: uterus; Vector: pT73-Pac; Site:1: Not I -
Site:2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
AACTGAGAGATTGGCGCCGCTTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="470954"
/clone_lib="Soares pregnant uterus NBHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
complement(<1..>501)
/db_xref="GDB:3756956"

BASE COUNT 141 a 122 c 87 g 146 t 5 others
ORIGIN

mrna

Query Match 100.0%; Score 87; DB 6; Length 501;
Best Local Similarity 98.9%; Pred. No. 3,21e-74;
Matches 87; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 5 TTGAGTGTGAGATGATTTATTAGACAAATTACATTTACTATTTTCTT 64

Cp 86 TTGAGTGTGAGATGATTTATTAGACAAATTACATTTACTATTTTCTT 29

Db 65 CTAACCCCTCACCATCAAGATTGATC 92

Cp 28 CTAACCCCTCACCATCAAGATTGATC 1

RESULT 7 AA635684 324 bp mrna EST 31-OCT-1997

LOCUS nrl3c01.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1169760,

DEFINITION mRNA sequence.

ACCESSION AA635684

KEYWORDS g2559526

EST.

human.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 324)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1350

Email: Robert_Strausberg@nih.gov

Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 402 Std Error: 0.00

Seq primer: -40M13 fwd. ET from Amerisham.

Location/Qualifiers

1. 324

/organism="Homo sapiens"

/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)

with a modified polylinker; 1st strand cDNA was prepared
from normal prostate bulk tissue, and was then primed with
a Not I - Oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library is normalized, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/db_xref="IMAGE:1169760"
/clone="IMAGE:1169760"
/clone_lib="NCI_CGAP_Pr22"
/sex="male"
/tissue_type="normal prostate"
/lab_host="DH10B"

BASE COUNT 107 a 76 c 50 g 91 t
ORIGIN

Query Match 97.7%; Score 85; DB 10; Length 324;
Best Local Similarity 98.8%; Pred. No. 8,90e-72;
Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 TAGATGTAGATAGATTTTATTAGACAAATTACATTTACTATTTTCTTCT 60

Cp 86 TAGATGTAGATAGATTTTATTAGACAAATTACATTTACTATTTTCTTCT 27

Db 61 AAACCCCTCACCATCAAGATTGATC 86

Cp 26 AAACCCCTCACCATCAAGATTGATC 1

RESULT 8 N27289 385 bp mrna EST 29-DEC-1995

LOCUS yw71b11.s1 Homo sapiens cDNA clone 257661 3',

DEFINITION N27289

ACCESSION g1141637

NID

KEYWORDS

EST.

human

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata;

Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;

Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;

Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 385)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, B., Rohlfing, T., Soares, M., Tan, F.,

Trevasalis, E., Waterston, R., Williamson, A., Wohlmann, P. and

Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

JOURNAL

COMMENT

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

High quality sequence stops: 305

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Location/Qualifiers

1. 385

/organism="Homo sapiens"

/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)

```

/organism="Homo sapiens"
/clone="257661"
BASE COUNT      115 a      96 c      65 g      109 t
ORIGIN
Query Match      97.7%  Score 85; DB 19; Length 385;
Best Local Similarity 98.8%; Pred. No. 8.90e-72;
Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 TAGATGATAGATAGATTTTATTAGACAATTTACATTTTACTATATTTTCCTTCT 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 86 TAGATGTANAGATAGATTTTATTAGACAATTTACATTTTACTATATTTTCCTTCT 27
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 61 AAACCTCACCATCAAGATTGATC 86
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 26 AAACCTCACCATCAAGATTGATC 1
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 9
LOCUS      AA043991      416 bp      mRNA      EST      10-MAY-1997
DEFINITION 2K5H02.s1 Soares pregnant uterus NBHPU Homo sapiens cDNA clone
ACCESSION  AA043991
NID        91521849
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae;
            Homo.
REFERENCE  1 (bases 1 to 416)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevasakis,E.,
            Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
            Washu-Merck EST Project
            Unpublished (1995)

TITLE
JOURNAL
COMMENT
Contact: Wilson RK
Washu-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 772 Std. Error: 0.00
Seq. primer: -40M13 fwd. from Amersham
High quality sequence stop: 370.
Location/Qualifiers
1. 416
/organism="Homo sapiens"
/notes="Organ: uterus; Vector: pRT73-Pac; Site.1: Not I;
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
AACTGGAGAGATTCGGCGCCGCTTTTCTTTTCTTTT 3'] ,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="487059"
/clone.lib="Soares pregnant uterus NBHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
complement(<1..>416)
/db_xref="GDB:3760773"

BASE COUNT      124 a      103 c      71 g      117 t      1 others
ORIGIN

```

```

Query Match      96.6%  Score 84; DB 6; Length 416;
Best Local Similarity 98.8%; Pred. No. 1.47e-70;
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 AGATGTAGATAGATTTTATTAGACAATTTACATTTTACTATATTTTCCTTCTA 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 85 AGATGTANAGATAGATTTTATTAGACAATTTACATTTTACTATATTTTCCTTCTA 26
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 61 AAACCTCACCATCAAGATTGATC 85
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 25 AAACCTCACCATCAAGATTGATC 1
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 10
LOCUS      H62857      422 bp      mRNA      EST      10-OCT-1995
DEFINITION YF18B08.s1 Homo sapiens cDNA clone 205623 3'.
ACCESSION  H62857
NID        91017203
KEYWORDS   EST.
SOURCE     human clone=205623 primer-Promega -21m13 library-Soares fetal liver
            spleen INF15 vector-pRT73D (Pharmacia) with a modified polylinker
            host-DH10B (ampicillin resistant) Raitel-Pac I Raitel-Eco RI liver
            and spleen from a 20 week post conception male fetus. 1st strand
            cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAGAGATTTTAAAGATCTTTTCTTTTCTTTT 3'] , double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pRT73
            vector. Library went through one round of normalization. Library
            constructed by Bento Soares and M. Fatima Bonaldo.

ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 422)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
            Wilson,R.
            Washu-Merck EST Project
            Unpublished (1995)

TITLE
JOURNAL
COMMENT
Contact: Wilson RK
Washu-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
High quality sequence stops: 279
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Location/Qualifiers
1. 422
/organism="Homo sapiens"
/clone="205623"

FEATURES
source
1. 422
/organism="Homo sapiens"
/clone="205623"

BASE COUNT      117 a      99 c      86 g      117 t      3 others
ORIGIN
Query Match      96.6%  Score 84; DB 19; Length 422;
Best Local Similarity 98.8%; Pred. No. 1.47e-70;
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 AGATGTAGATAGATTTTATTAGACAATTTACATTTTACTATATTTTCCTTCTA 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 85 AGATGTANAGATAGATTTTATTAGACAATTTACATTTTACTATATTTTCCTTCTA 26
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 61 AAACCTCACCATCAAGATTGATC 85
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

Cp 25 AAACCTCACCATCAAGATTGATC 1

RESULT 11
LOCUS AA700634 562 bp mRNA EST 19-DEC-1997
DEFINITION Z143B04.s1 Soares fetal liver spleen INFUS S1 Homo sapiens cDNA
ACCESSION AA700634
KEYWORDS Clone 433519 3', mRNA sequence.
SOURCE EST
ORGANISM human.
REFERENCE
AUTHORS

1 (bases 1 to 562)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kritzman, D., Kucabada, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
Martini, J., Moore, B., Schellenberg, K., Stepien, M., Tan, F.,
Theriot, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
COMMENT
TITLE
JOURNAL
AUTHORS

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m3 fwd. Et from Amersham
High quality sequence stop: 461.
Location/Qualifiers

1. 562
/organism="Homo sapiens"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subcloned version of the original Soares fetal
liver spleen INFUS library. 1st strand cDNA was primed
with a Pac I - 01190(dT) primer (5',
AATCGAAGATTAATTAAGATCTTTTCTTTTCTTTT 3'),
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="GB:133288"
/db_xref="taxon:9606"
/clone_lib="Soares fetal liver spleen INFUS S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT 164 a 145 c 93 g 159 t 1 others
ORIGIN

Query Match 96.6%; Score 84; DB 11; Length 562;
Best Local Similarity 98.8%; Pred. No. 1.47e-70;
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 AGATGTAGATAGATTATTTAAGACAATTACACATTTTCTATATTTCCCTTCA 60
|||||
Cp 85 AGATGTAGATAGATTATTTAAGACAATTACACATTTTCTATATTTCCCTTCA 26
Db 61 AAACCTCACCATCAAGATTGATC 85
|||||
Cp 25 AAACCTCACCATCAAGATTGATC 1

RESULT 12
LOCUS AA347704 414 bp mRNA EST 21-APR-1997
DEFINITION EST54313 Fetal heart II Homo sapiens cDNA 3' end, mRNA sequence.
ACCESSION AA347704

NID 92000170
KEYWORDS
EST.
SOURCE
ORGANISM

Homo sapiens
Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae;
Homo.
1 (bases 1 to 414)
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,
Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,
White, O., Sutton, G., Blake, U.A., Brandon, R.C., Man, Wai, C.,
Fitzgerald, L.M., Fitch, W.M., Colton, M.D., Earle-Hughes, J., Fine, L.D.,
Glocke, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,
Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, L.A., Sauder, D.M., Shirley, R.,
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
Bednarek, D.P., Gao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
Dimke, D., Feng, D.-F., Ferrile, A., Fischer, C., Hastings, G.A.,
He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,
Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meisner, P.S., Olsen, H.,
Raymond, L., Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruden, S.M.,
Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,
Fraser, C.M. and Venter, J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl.), 3-174 (1995)
96026280
Other ESTs: EST54314 THC180445
CONTACT: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerl@vtrig.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi.html>)
Seq primer: M13-21.
Location/Qualifiers

1. 414
/organism="Homo sapiens"
/note="Organ: heart; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
/db_xref="ATCC (dbost):149009"
/db_xref="taxon:9606"
/clone_lib="Fetal heart II"
/dev_stage="Fetus"
complement(<1..>414)
BASE COUNT 121 a 102 c 70 g 113 t 8 others
ORIGIN

Query Match 94.3%; Score 82; DB 25; Length 414;
Best Local Similarity 96.5%; Pred. No. 3.99e-68;
Matches 82; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 1 AGATGTAGATAGATTATTTAAGACAATTACACATTTTCTATATTTCCCTTCA 60
|||||
Cp 85 AGATGTAGATAGATTATTTAAGACAATTACACATTTTCTATATTTCCCTTCA 26
Db 61 AAACCTCACCATCAAGATTGATC 85
|||||
Cp 25 AAACCTCACCATCAAGATTGATC 1

RESULT 13
LOCUS AA928824 317 bp mRNA EST 23-JUN-1998
DEFINITION on98c05.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1564808 3', mRNA sequence.
ACCESSION AA928824
NID 93078181

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 317)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 810 Std Error: 0.00
Seq primer: -40m13 fwd. ET from AmerSham
High quality sequence stop: 268.
Location/Qualifiers
1. 317
/organism="Homo sapiens"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker. Site.1: Not I; Site.2: Eco RI.
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCL-CGAP-GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682633-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bernaldo."
/db_xref="taxon:9606"
/db_xref="IMAGE:1564808"
/clone_1lb="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"

BASE COUNT 105 a 77 c 45 g 89 t 1 others
ORIGIN

Query Match 89.7%; Score 78; DB 15; Length 317;
Best Local Similarity 97.7%; Pred. NO. 2.78e-63;
Matches 84; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 3 AGATGAGATGATGATTTTATTAAGACACATTCATTTCTATTTCTTCTTCT 62
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 85 AGATGAGATGATGATTTTATTAAGACACATTCATTTCTATTTCTTCTTCT 27
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 63 AAACCCCTCACCATCAAGATTGATC 88
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 26 AAACCCCTCACCATCAAGATTGATC 1
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

LOCUS 14 T57478 307 bp mRNA EST 08-FEB-1995
DEFINITION yb56d02.s1 Homo sapiens cDNA clone 75171 3'.
ACCESSION T57478
NID 9659339
KEYWORDS EST.
SOURCE human clone-75171 library-Stratagene ovary (#937217)
vector-Blaescript SR host-SOLR cells (kanamycin resistant)
primer-21m13 Rsite1-EcoRI Rsite2-XhoI Cloned unidirectionally.
Primer: 01m13 dt. Total ovary tissue, normal 49 year old caucasian
female. Average insert size: 0.8 kb; Uni-ZAP XR Vector; 5' adaptor
sequence: 5'-GAATTCGGCAGAG-3'; 3' adaptor sequence:
5'-CTCCAGTTTCTTTTCTTTTCTTTT-3'.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 307)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Matra, M.,

TITLE
JOURNAL
COMMENT
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
High quality sequence stops: 280
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Location/Qualifiers
1. 307
/organism="Homo sapiens"
/clone="75171"

BASE COUNT 102 a 75 c 45 g 84 t 1 others
ORIGIN

Query Match 87.4%; Score 76; DB 5; Length 307;
Best Local Similarity 96.7%; Pred. NO. 7.15e-61;
Matches 76; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 AGATGAGATGATGATTTTATTAAGACACATTCATTTCTATTTCTTCTTCTTCT 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 77 AGATGAGATGATGATTTTATTAAGACACATTCATTTCTATTTCTTCTTCTTCT 18
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 61 ACCATCAAGATTGATC 77
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 17 ACCATCAAGATTGATC 1
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 15
LOCUS AA862584 452 bp mRNA EST 24-AUG-1998
DEFINITION oh44e12.s1 NCI-CGAP_GC4 Homo sapiens cDNA clone IMAGE:1469518 3',
mRNA sequence.
ACCESSION AA862584
NID 92955063
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 452)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL.at:
www.bio.llnl.gov/bdpr/image/image.html
Insert Length: 1964 Std Error: 0.00
Seq primer: -40m13 fwd. ET from AmerSham
High quality sequence stop: 431.
Location/Qualifiers
1. 452
/organism="Homo sapiens"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled

[REDACTED]
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_un n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Tue Dec 22 05:32:37 1998: Maspar time 30.59 Seconds
Tabular output not generated. 391.360 Million cell updates/sec

Title: >US-08-530-112A-3084
Description: (1-88) from US08530112A.seq
Perfect Score: 87
N.A. Sequence: 1 GATCAATCTTTGATGCTGAG.....AATCTACTTACATCTTAA 88
Comp: CTACTGTAACACTACACTC.....TTAGATGANAATGTAGATTT

Scoring table: TABLE default
Gap 6

Nmatch SMD : Dbase 0; Query 0

Searched: 188442 segs, 68026449 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 6.701; Variance 5.314; scale 1.261

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	87	100.0	88	19	T22084		Human gene signature	1.14e-31
2	30	34.5	91	9	Q51746		Oligonucleotide probe	7.79e-04
3	28	33.2	204	1	N81164		Base substituted E.co	5.90e-03
4	27	31.0	91	9	O51746		Oligonucleotide probe	1.60e-02
5	25	28.7	204	1	N81164		Base substituted E.co	1.15e-01
6	24	27.6	29544	17	T12460		Ovine adenovirus geno	3.02e-01
7	22	25.3	67	24	T14322		Primer used in the la	2.02e+00
8	22	25.3	172	22	T26787		Human gene signature	2.02e+00
9	22	25.3	2208	30	T66533		Tomato E8 promoter.	2.02e+00
10	22	25.3	2216	20	T09974		E8 promoter.	2.02e+00
11	22	25.3	2216	13	Q78007		Tomato E8 gene promot	2.02e+00
12	22	25.3	2216	17	T09965		Tomato E8 promoter re	2.02e+00
13	22	25.3	7635	15	Q84654		Human neuronal calciu	2.02e+00

c	14	22	25.3	7635	6	Q37811	Sequence encoding the	2.02e+00
c	15	22	25.3	7635	15	Q84653	Human neuronal calciu	2.02e+00
c	16	22	25.3	580073	27	T58840	Mycoplasma genitalium	2.02e+00
c	17	22	24.1	68	34	T63255	Messenger RNA primer	5.13e+00
c	18	21	24.1	68	34	T73397	Oligonucleotide tag c	5.13e+00
c	19	21	24.1	70	24	T14325	Conjugate formed by 1	5.13e+00
c	20	21	24.1	501	3	N50033	Sequence encoding new	1.28e+01
c	21	20	23.0	705	30	T72057	Sequence flanking mar	1.28e+01
c	22	20	23.0	705	30	T43937	Sequence flanking mar	1.28e+01
c	23	20	23.0	1045	30	T72058	Sequence flanking mar	1.28e+01
c	24	20	23.0	1045	30	T43938	Sequence flanking mar	1.28e+01
c	25	20	23.0	1474	38	T90174	Oil seed rape cystein	1.28e+01
c	26	20	23.0	1578	32	T79597	TARA-binding protein	1.28e+01
c	27	20	23.0	1578	25	T42223	Human TARA-binding pr	1.28e+01
c	28	20	23.0	2504	31	T73286	S. pombe origin of re	1.28e+01
c	29	20	23.0	2504	30	T62359	Schizosaccharomyces p	1.28e+01
c	30	19	21.8	39	24	T14323	Sequence used in the	3.14e+01
c	31	19	21.8	109	2	N60058	Transcriptional regul	3.14e+01
c	32	19	21.8	189	22	T28476	Human gene signature	3.14e+01
c	33	19	21.8	501	3	N50029	Sequence encoding new	3.14e+01
c	34	19	21.8	501	3	N50023	Sequence encoding new	3.14e+01
c	35	19	21.8	501	3	N50030	Sequence encoding new	3.14e+01
c	36	19	21.8	501	3	N50024	Sequence encoding new	3.14e+01
c	37	19	21.8	501	3	N50032	Sequence encoding new	3.14e+01
c	38	19	21.8	1408	22	T28260	Survival motor neuron	3.14e+01
c	39	19	21.8	1491	22	T31100	Papillomavirus E2 bin	3.14e+01
c	40	19	21.8	1582	22	T28259	Survival motor neuron	3.14e+01
c	41	19	21.8	1582	22	T18831	Human survival motor	3.14e+01
c	42	19	21.8	2223	13	Q80908	Plasmodium falciparum	3.14e+01
c	43	19	21.8	3273	22	T18868	Human survival motor	3.14e+01
c	44	19	21.8	3600	36	T77330	Solanum tuberosum col	3.14e+01
c	45	19	21.8	9636	12	Q67190	P. falciparum transmi	3.14e+01

ALIGNMENTS

RESULT 1
ID T22084 standard; cDNA to mRNA; 88 BP.

AC T22084;
DT 02-AUG-1996 (first entry)

DE Human gene signature HUMGS03628.

KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;

KW human; cloning; mapping; non-biased library; diagnosis; detection;

KW cell typing; abnormal cell function; ss.

OS Homo sapiens.

PN WO9514772-A1.

PD 01-JUN-1995.

PF 11-NOV-1994; J01916

PR 12-NOV-1993; JF-355504.

PA (MATS/) MATSUBARA K.

PI (OKUB/) OKUBO K.

PI Matsubara K, Okubo K;

DR WPI: 95-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

PT tissues

PS Claim 1: Page 1038; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in T19001-T26837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

```

PS Disclosure; p: English
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are misincorporated by the
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 32.28; Score 28; DB 1; Length 204;
Best Local Similarity 8.38; Pred. No. 5,90e-03;
Matches 6; Conservative 43; Mismatches 21; Indels 2; Gaps 2

Db 98 rbhvydrnsdaaawccyrsvkdcyachdhdyvbyvnhnnccbn 157
: : : : : : : : : : : : : : : : : : : : : : : : :
Cp 71 ATTATTATGACAAATATCA-CATTTCACATATTTCCTTCTAAACC-CTCACCA 14
: : : : : : : : : : : : : : : : : : : : : : : : :
Db 158 nhvchnvbnhn 169
: : : : : : : : : : : : : : : : : : : : : : : : :
Cp 13 TCAACAGTTGAT 2

RESULT 4
ID 051746 standard; cDNA; 91 BP.
AC 051746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KM Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
SN Synthetic.
PS EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI: 93-378844/46.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3: Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (051735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also 051735-45 and 051747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 31.08; Score 27; DB 9; Length 91;
Best Local Similarity 0.08; Pred. No. 1.60e-02;
Matches 0; Conservative 35; Mismatches 8; Indels 0; Gaps 0;

Db 16 yvvvshhsvhvnhvsvvvhvnhvhyvhyv 58
: : : : : : : : : : : : : : : : : : : : : : : : :
Cp 10 TTGATGCTGAGGGTTTACAAAGAAATAATAGTAAATGCG 52
: : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
ID NB1164 standard; DNA; 204 BP.
AC NB1164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KM E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key
TT Location/Qualifiers
misc_feature 19..69
/att=

```



```

FT      primer_bind      /function=multiple cloning site
FT      187..204
FT      /*tag= b
PD      EP-285123-A.
PD      05-MAY-1988.
PD      30-MAR-1988. 105163.
PD      03-APR-1987: US-034819.
PD      (SUSO) SUOMEN SOKERI OY.
PI      Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
PI      WPI: 86-27927/40.
PI      Introducing random point mutations into nucleic acids -
PI      by prepn of single stranded template, annealing a primer, elongation,
PI      misincorporation, completion of molecules and screening.
PS      Disclosure: P. English.
CC      Random point mutations were introduced into the alpha fragment of
CC      E. coli beta-galactosidase. The wild type sequence was obtained as a
CC      single stranded template and an oligonucleotide was hybridised to
CC      it to generate a popn of DNA molecules which terminate at all
CC      possible nucleotide positions within a specified region. The
CC      variable 3' ends generated in this way are used as primers for
CC      reverse transcriptase. Nucleotides are misincorporated by the
CC      transcriptase and the molecules are completed to forms that can be
CC      amplified and then expressed in a suitable host-vector system.
CC      The sequence covers all 176 diff base substitutions, most of which
CC      occurred singularly in any given mutant.
CC      See also P80575.
SC      Sequence 204 BP. 21 A; 47 C; 17 G; 11 T; 108 Others;
SO

Query Match      28.7%; Score 25; DB 1; Length 204;
Best Local Similarity 12.9%; Pred. No. 1.15e-01;
Matches      8; Conservative 35; Mismatches 18; Indels 1; Gaps 1;

DB      88 rttchyrirnbvdyrdyn-rsdaaayccyrrsvkydcynachdhdyvbybvnvh 146
OY      22 GTTATGAGAACGAAAAATAGTAAATGTGTAATTGTCTAATATAAATCATCTATAC 81
DB      147 nh 148
OY      82 AT 83

```

```

DR      WPI: 96-117053/12.
PT      Viral vector derived from ovine adenovirus (OAV287) - used for
PT      recombinant prodn. and delivery of non-adenoviral polypeptide(s) or
PT      RNA molecules in host animals, esp. sheep.
PS      Claim 1; Fig 1; 53pp; English.
CC      The sequence represents the entire genome of ovine adenovirus OAV287,
CC      a benign adenovirus isolated from sheep in Australia, and begins at
CC      base 1 of the left-hand inverted terminal repeat. The sequence has
CC      been determined by cloning BamHI restriction fragments representing
CC      the entire genome and sequencing nested deletion mutants. Reading
CC      frames have been tentatively identified, including E1A/B, E4 and E3.
CC      The sequence contains promoters, inverted terminal repeats, viral
CC      packaging signals and RNA processing signals. The entire genomic
CC      clone is present in plasmid pOAV100. It is possible to introduce
CC      recombinant DNA at non-essential sites, e.g. in the E3 region,
CC      including a rotavirus VP7sc antigen, Trichostongyrius colubriformis
CC      17-kDa antigen, Taenia ovis 45W antigen or Lucilia cuprina p95
CC      antigen gene, or DNA encoding antisense RNA or a ribozyme. The
CC      sequence may be used in construction of vectors to express
CC      heterologous DNA in a grazing animal, e.g. a sheep. Many host
CC      species have no pre-existing immunity to the virus, allowing
CC      efficient infection, gene delivery and expression, and the
CC      possibility of virus spread to a non-target host is minimized. The
CC      virus may be used as a vector for recombinant vaccine production or
CC      transgenic animal construction.
SC      Sequence 29544 BP. 10167 A; 5365 C; 4562 G; 9450 T;
SO

Query Match      27.6%; Score 24; DB 17; Length 29544;
Best Local Similarity 81.6%; Pred. No. 3.02e-01;
Matches      31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB      27029 atgtatataacacatttcatactctt 27066
CP      71 AATTATATAGACAAATACACATTTTACTATATTTT 34

```

```

RESULT 7
ID      T14322 standard; DNA: 67 BP.
AC      T14322;
DT      16-JAN-1997 (first entry)
DE      Primer used in the labelling and sorting of nucleotide molecules.
KW      Labelling; sorting; sequencing; tag; tagging; ss.
OS      Synthetic.
FH      Key
FH      misc_feature 67
FT      Location/Qualifiers
FT      /*tag= a
FT      /*mod_base= Linked to biotin
PD      MO9612039-A1.
PD      23-APR-1996.
PD      12-OCT-1995: 012678.
PD      13-OCT-1994: US-322348.
PD      19-DEC-1994: US-359295.
PD      (LYNX-) LYNX THERAPEUTICS INC.
PI      Brenner S;
PI      WPI: 96-222023/22.
PI      Labelling and sorting mols. using oligo:nucleotide tags - useful in
PI      large-scale parallel operations, e.g. DNA sequencing and mRNA
PI      fingerprinting
PS      Disclosure: Page 19; 71pp; English.
CC      Determining the nucleotide sequence (1) of a target polynucleotide
CC      (T) comprises: (a) generating from T a plurality of fragments that
CC      cover T; (b) attaching an oligonucleotide tag from a repertoire of
CC      tags, to each fragment such that all the same fragments have the
CC      same tag, and all different fragments have different tags; (c)
CC      sorting the fragments by specifically hybridising the tags with
CC      their respective tag complements; (d) determining (1) of a portion
CC      of each of the fragments; and (e) determining (1) of T by collating
CC      the sequences of the fragments. The tagging system can be used with
CC      single base sequencing methods to sequence polynucleotides up to
CC      several kilobases in length. The tagging system permits many
CC      thousands of fragments of a target polynucleotide to be sorted onto
CC      one or more solid phase supports and sequenced simultaneously.
CC      This sequence represents a primer which could be used to reverse

```

```

RESULT 6
ID      T12460 standard; DNA: 29544 BP.
AC      T12460;
DT      21-APR-1996 (first entry)
DE      Ovine adenovirus genome.
KW      Ovine adenovirus; genome; sheep; inverted terminal repeat; E1A/B;
KW      E4; E3; promoter; packaging signal; RNA processing signal;
KW      plasmid pOAV100; cloning; vector; antigen; recombinant vaccine;
KW      rotavirus; Trichostongyrius colubriformis; Taenia ovis;
KW      Lucilia cuprina; antisense RNA; ribozyme; transgenic animal;
KW      gene transfer; cyclic; ss.
KW      Ovine adenovirus strain OAV287.
OS      Ovine adenovirus strain OAV287.
FH      Key
FH      repeat_region 1..46
FH      Location/Qualifiers
FH      /*tag= a
FH      /*rpl_type= INVERTED
FH      /*note= "Inverted terminal repeat (claim 7)"
PD      misc_feature 22139..22130
PD      /*tag= b
PD      /*note= "Non-essential region suitable for cloning"
PD      26645..26646
PD      /*tag= c
PD      /*note= "Non-essential region suitable for cloning"
PD      28644..28649
PD      /*tag= d
PD      /*note= "Non-essential region suitable for cloning"
PD      WO9603308-A1.
PD      08-FEB-1996.
PD      26-JUL-1995: AU0453.
PD      26-JUL-1994: AU-007101.
PD      (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PI      Both GW, Boyle DB, Vratil S;

```


Transgenic plants with modified ripening phenotype - contg. a chimeric gene contg. a sequence which encodes a prod. which reduces ethylene biosynthesis

PS Example 1: Page 90-92: 139pp: English.

CC The sequence represents a tomato E8 promoter, isolated as the upstream -2216 bp region of the tomato E8 gene. The promoter fragment is both stage- and tissue-specific in tomato, being induced at the onset of ripening and activated by ethylene. The lower E8 promoter fragment may be isolated using primers T09983-4, and the upper fragment may be isolated using this fragment as a DNA probe, followed by insertion upstream of the lower promoter in plasmid pAG-5321 at HindIII and XbaI sites, to give plasmid pGA-SESKN. A gene which modifies ripening characteristics, e.g. an S-adenosylmethionine-hydroxylase, aminocyclopropane-1-carboxylate-deaminase, aminocyclopropane-1-carboxylate-synthase or aminocyclopropane-1-carboxylate-oxidase gene or corresponding antisense sequence, may be inserted under the control of this promoter and introduced into e.g. tomato to produce a transgenic plant bearing fruit with altered fruit ripening phenotype, and a subsequent reduction in ethylene biosynthesis.

SO Sequence 2216 BP; 866 A; 253 C; 290 G; 805 T;

Query Match 25.3%; Score 22; DB 20; Length 2216;
Best Local Similarity 82.4%; Pred. No. 2.02e+00;
Matches 28; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 919 aaaaagaagaaatttcaatttttcaata 952
||||| ||||||| ||||||| ||||||| |||||||
OY 34 AAAAATATAGTAAATGCTAATTGCTTAATA 67

RESULT 11
ID 078007 standard; DNA: 2216 BP.
AC 078007;
DT 05-JUL-1995 (first entry)
DE Tomato E8 gene promoter.
KW Tomato E8 promoter; transgenic fruit; ethylene production inhibition;
KW S-adenosylmethionine hydroxylase; delayed fruit ripening; ds.
OS Lycopersicon esculentum.
PN W09424294-A.
PD 27-OCT-1994.
PE 08-APR-1994; US-046583.
PR 09-APR-1993; US-046583.
PA (EPIIT-) EPIITOP INC.
PI Bestwick RK, Ferro AJ;
DR WPI: 94-341873/42.
PT Transgenic fruit-bearing plants, esp. tomato - contg. exogenous gene under control of E8 promoter
PS Claim 3; Fig 13: 59pp: English.
CC 078007 the tomato E8 gene promoter, and 078008 which encodes R74824 the bacteriophage T3 derived SAM-K modified S-adenosylmethionine hydroxylase (AdoMetase), were used in the construction of a vector.
CC The vector was used to produce transgenic tomato plants which expressed AdoMetase under the translational control of the E8 promoter. AdoMetase is an inhibitor of ethylene production, and therefore an inhibitor of fruit ripening.
CC Sequence 2216 BP; 866 A; 253 C; 290 G; 805 T;

Query Match 25.3%; Score 22; DB 13; Length 2216;
Best Local Similarity 82.4%; Pred. No. 2.02e+00;
Matches 28; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 919 aaaaagaagaaatttcaatttttcaata 952
||||| ||||||| ||||||| ||||||| |||||||
OY 34 AAAAATATAGTAAATGCTAATTGCTTAATA 67

RESULT 12
ID T09965 standard; DNA: 2216 BP.
AC T09965;
DT 16-MAY-1996 (first entry)
DE Tomato E8 promoter region.

KW S-adenosylmethionine hydroxylase; AdoMet hydroxylase; AdoMetase; SAM-K; S-adenosyl cleaving enzyme; SAMase; raspberry; Rubus idaeus; Strawberry; Fragaria ananassa; transgenic plant; fruit rot; disease resistance; fungus resistance; virus resistance; ethylene; KW ripening; crop improvement; tomato; promoter E8; probe; ds.
OS Lycopersicon esculentum.
FH Key location/Qualifiers
FT misc_difference 227
FT /*tag- a
FT /note- "base n at position 227 is not identified
FT in the specification"
FT misc_difference 804
FT /*tag- b
FT /note- "base n at position 804 is not identified
FT in the specification"
PN W09535388-A1.
PD 28-DEC-1995.
PE 16-JUN-1995; U07753.
PR 17-JUN-1994; US-263900.
PR 03-FEB-1993; US-384356.
PA (EPIIT-) EPIITOP INC.
PI Bestwick RK, Ferro AJ, Mathews HV;
DR WPI: 96-058425/06.
PT Transgenic red raspberry and strawberry plants - produce fruit with increased sugar content, increased fungal or viral resistance, or reduced ethylene biosynthesis
PS Disclosure; Page 44-46; 85pp: English.
CC The regulatable tomato E8 gene promoter (T09965) can be used to control expression of foreign genes introduced into transgenic plants. It can also be used as a hybridisation probe. Transgenic CC raspberry and strawberry plants have been obtd. that produce fruit with increased sugar content, improved disease resistance, or reduced ethylene biosynthesis.
CC Sequence 2216 BP; 866 A; 253 C; 290 G; 805 T;

Query Match 25.3%; Score 22; DB 17; Length 2216;
Best Local Similarity 82.4%; Pred. No. 2.02e+00;
Matches 28; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 919 aaaaagaagaaatttcaatttttcaata 952
||||| ||||||| ||||||| ||||||| |||||||
OY 34 AAAAATATAGTAAATGCTAATTGCTTAATA 67

RESULT 13
ID 084654 standard; DNA: 7635 BP.
AC 084654;
DT 30-NOV-1995 (first entry)
DE Human neuronal calcium channel subunit alpha 1D including alternative exon encoding the I56 transmembrane domain.
KW Calcium channel subunit; antagonist; agonist; diagnosis;
KW Lambert Eaton Syndrome; ss.
OS Homo sapiens.
FH Key location/Qualifiers
FT cds 511..6996
FT /*tag- a
FT variation 1627..1728
FT /*tag- b
FT /label- alternative exon
PN W09504832-A.
PD 16-FEB-1995.
PE 11-AUG-1994; U09230.
PR 11-AUG-1993; US-105536.
PR 05-NOV-1993; US-149097.
PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;
DR WPI: 95-090900/12.
DR P-PSDB; R71002.
PT DNA encoding human calcium channel sub-unit(s) - used for obtaining agonists and antagonists
PS Disclosure; Page 126-127; 285pp: English.
CC The alpha 1D subunit cDNA has been isolated using fragments of the

```

ID RESULT 15
AC 084653 standard; cDNA; 7635 BP.
DE 084653:
DE 30-NOV-1995 (first entry)
DE Human neuronal calcium channel subunit alpha 1D.
DE Calcium channel subunit; antagonist; agonist; diagnosis;
KW Lambert Eaton Syndrome; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 511..6996
FT /tag= a
PN MO9504822-A.
PD 16-FEB-1995. U09230.
PF 11-AUG-1994; U09230.
PR 11-AUG-1993; US-105536.
PR 05-NOV-1993; US-149097.
PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
PI Ellis SB, Gillespie A, Harpold KM, McCue AF, Williams ME;
DR MPI: 95-090900/12.
DR P-PSDB; R71001.
PT DNA encoding human calcium channel sub-unit(s) - used for
PT developing prods. for studying calcium channels, e.g. for
PT obtaining agonists and antagonists
PS Disclosure; page 116-126; 283pp. English.
CC The alpha 1D subunit cDNA has been isolated using fragments of the
CC rabbit skeletal muscle calcium channel alpha 1 subunit cDNA as a
CC probe to screen a cDNA library of human neuroblastoma cell line
CC IMR32, to obtain clone alpha1.36. This clone was used as a probe to
CC screen additional IMR32 cell cDNA libraries to obtain overlapping
CC clones, which were then employed for screening until a sufficient
CC series of clones to span the length of the nt sequence encoding the
CC human alpha 1D subunit was obt'd. Full-length clones were then
CC constructed by ligating partial clones. Q84653 shows the nt sequence
CC of the cDNA encoding the alpha 1D subunit. The Alpha 1D protein has
CC a calculated Mr of 245,163. It contains four putative transmembrane
CC segments. It mediates DHP-sensitive high-voltage, long-lasting
CC calcium channel activity.
SQ Sequence 7635 BP; 1983 A; 1924 C; 1697 G; 1831 T;

Query Match 25.3%; Score 22; DB 15; Length 7635;
Best Local Similarity 72.3%; Pred. No. 2, 02e+00;
Matches 34; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Db 218 tacatgtatataattcattagaataatatacatcggatttattttt 264
|| | | | | | | | | | | | | | | | | | | | |
Cp 80 TANAGATAGATTATTTATTAAGACAATAATACATTTTACTATATTTT 34

Search completed: Tue Dec 22 05:33:13 1998
Job time : 36 secs.

```

[W] [E] [S] [E] [H]

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_n n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Tue Dec 22 05:25:48 1998; MasPar time 181.06 Seconds
1071.078 Million cell updates/sec
Tabular output not generated.

Title: >US-08-530-112A-3084
Description: (1-88) from US08530112A.seq
Perfect Score: 87
N.A. Sequence: 1 GATCAATCTTGTGATGGTGAG.....ATCATCTTACATCTTAA 88
Comp: CTAGTGAACCTACCACTC.....TTAGTAGANATGTAGATT

Scoring table: TABLE default
Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 567134 segs, 1101898692 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

emb155
1:em_ba 2:em_fun 3:em_htg 4:em_hum1 5:em_hum2 6:em_in
7:em_cm 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_vi
Database: genbank107
15:gb_ba1 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_cm 20:gb_ov
21:gb_pat 22:gb_ph 23:gb_pl1 24:gb_pl2 25:gb_pl
26:gb_pr2 27:gb_pr3 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
32:gb_un 33:gb_vi

Statistics: Mean 8.815; Variance 6.105; scale 1.444

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description	Pred. No.
C	1	76	87.4	307	30	G05576		human STS WI-8386.	9.74e-27
	2	37	42.5	2833	28	AF022136		Rattus norvegicus conn	6.87e-06
	3	37	42.5	3115	28	AF021806		Rattus norvegicus conn	6.87e-06
	4	30	34.5	158	19	AF028610		Capra hircus connexin4	1.53e-02
	5	29	33.3	27856	23	AB009051		Arabidopsis thaliana g	4.42e-02
	6	28	33.2	10772	18	AF012089		Drosophila melanogaste	1.26e-01
	7	28	32.2	83675	27	AC004830		Homo sapiens PAC clone	1.26e-01
	8	27	31.0	50641	23	AB008270		Arabidopsis thaliana g	3.53e-01
	9	27	31.0	128330	17	HS11106		Human DNA sequence ***	3.53e-01
	10	26	29.9	10772	18	AF012089		Drosophila melanogaste	9.75e-01
	11	25	28.7	311	18	HHU32460		Herneuphychia hermes N	2.65e+00
	12	25	28.7	2606	23	YSTM7GMX		Saccharomyces cerevisi	2.65e+00
	13	25	28.7	2606	23	MISCTRNM		yeast mitochondrial ge	2.65e+00

14	25	28.7	3980	23	MISC37	Yeast mitochondrial ge	2.65e+00
15	25	28.7	4350	23 <td>YSCMTG11</td> <td>Saccharomyces cerevisi</td> <td>2.65e+00</td>	YSCMTG11	Saccharomyces cerevisi	2.65e+00
16	25	28.7	43927	25 <td>HS0244G10</td> <td>Human DNA sequence fro</td> <td>2.65e+00</td>	HS0244G10	Human DNA sequence fro	2.65e+00
17	25	28.7	71736	23 <td>AB009048</td> <td>Arabidopsis thaliana g</td> <td>2.65e+00</td>	AB009048	Arabidopsis thaliana g	2.65e+00
18	24	27.6	1448	18 <td>AF004169</td> <td>Apis mellifera ultravil</td> <td>7.09e+00</td>	AF004169	Apis mellifera ultravil	7.09e+00
19	24	27.6	5724	33 <td>OAU40838</td> <td>Ovine adenovirus e4 30</td> <td>7.09e+00</td>	OAU40838	Ovine adenovirus e4 30	7.09e+00
20	24	27.6	21302	18 <td>PSC004088</td> <td>Plasmodium falciparum</td> <td>7.09e+00</td>	PSC004088	Plasmodium falciparum	7.09e+00
21	24	27.6	81850	23 <td>ATPC49</td> <td>Arabidopsis thaliana D</td> <td>7.09e+00</td>	ATPC49	Arabidopsis thaliana D	7.09e+00
22	24	27.6	207674	23 <td>ATPC48</td> <td>Arabidopsis thaliana D</td> <td>7.09e+00</td>	ATPC48	Arabidopsis thaliana D	7.09e+00
23	24	27.6	240114	17 <td>HS90L6</td> <td>Human DNA sequence ***</td> <td>7.09e+00</td>	HS90L6	Human DNA sequence ***	7.09e+00
24	23	26.4	382	30 <td>HS238YC11</td> <td>H. sapiens DNA segment</td> <td>1.88e+01</td>	HS238YC11	H. sapiens DNA segment	1.88e+01
25	23	26.4	690	18 <td>DMGAT6</td> <td>D. melanogaster DNA, GA</td> <td>1.88e+01</td>	DMGAT6	D. melanogaster DNA, GA	1.88e+01
26	23	26.4	870	28 <td>MUSMAB21</td> <td>Mouse hMC class II h2-</td> <td>1.88e+01</td>	MUSMAB21	Mouse hMC class II h2-	1.88e+01
27	23	26.4	12548	26 <td>AF076974</td> <td>Homo sapiens TRAP pro</td> <td>1.88e+01</td>	AF076974	Homo sapiens TRAP pro	1.88e+01
28	23	26.4	26291	18 <td>CELC317</td> <td>Caenorhabditis elegans</td> <td>1.88e+01</td>	CELC317	Caenorhabditis elegans	1.88e+01
29	23	26.4	30101	18 <td>PSC03015</td> <td>Plasmodium falciparum</td> <td>1.88e+01</td>	PSC03015	Plasmodium falciparum	1.88e+01
30	23	26.4	31757	27 <td>HSN53A9</td> <td>Human DNA sequence fro</td> <td>1.88e+01</td>	HSN53A9	Human DNA sequence fro	1.88e+01
31	23	26.4	33268	18 <td>CELF08D12</td> <td>Caenorhabditis elegans</td> <td>1.88e+01</td>	CELF08D12	Caenorhabditis elegans	1.88e+01
32	23	26.4	38955	18 <td>CEZC84</td> <td>Caenorhabditis elegans</td> <td>1.88e+01</td>	CEZC84	Caenorhabditis elegans	1.88e+01
33	23	26.4	45811	24 <td>AB017070</td> <td>Arabidopsis thaliana g</td> <td>1.88e+01</td>	AB017070	Arabidopsis thaliana g	1.88e+01
34	23	26.4	66084	24 <td>AB015468</td> <td>Arabidopsis thaliana g</td> <td>1.88e+01</td>	AB015468	Arabidopsis thaliana g	1.88e+01
35	23	26.4	86098	24 <td>ATP13D4</td> <td>Arabidopsis thaliana D</td> <td>1.88e+01</td>	ATP13D4	Arabidopsis thaliana D	1.88e+01
36	23	26.4	90019	23 <td>TM021B04</td> <td>Arabidopsis thaliana B</td> <td>1.88e+01</td>	TM021B04	Arabidopsis thaliana B	1.88e+01
37	23	26.4	91318	24 <td>ATAC005313</td> <td>Arabidopsis thaliana C</td> <td>1.88e+01</td>	ATAC005313	Arabidopsis thaliana C	1.88e+01
38	23	26.4	96792	26 <td>HS095741</td> <td>Human chromosome 16p13</td> <td>1.88e+01</td>	HS095741	Human chromosome 16p13	1.88e+01
39	23	26.4	97789	23 <td>AC004255</td> <td>Human chromosome 16p13</td> <td>1.88e+01</td>	AC004255	Human chromosome 16p13	1.88e+01
40	23	26.4	100754	17 <td>AC005489</td> <td>*** SEQUENCING IN PROG</td> <td>1.88e+01</td>	AC005489	*** SEQUENCING IN PROG	1.88e+01
41	23	26.4	110000	17 <td>CEY80D3</td> <td>Caenorhabditis elegans</td> <td>1.88e+01</td>	CEY80D3	Caenorhabditis elegans	1.88e+01
42	23	26.4	114058	17 <td>AC005009</td> <td>*** SEQUENCING IN PROG</td> <td>1.88e+01</td>	AC005009	*** SEQUENCING IN PROG	1.88e+01
43	23	26.4	153019	17 <td>AC004969</td> <td>*** SEQUENCING IN PROG</td> <td>1.88e+01</td>	AC004969	*** SEQUENCING IN PROG	1.88e+01
44	23	26.4	160112	17 <td>AC005102</td> <td>*** SEQUENCING IN PROG</td> <td>1.88e+01</td>	AC005102	*** SEQUENCING IN PROG	1.88e+01
45	23	26.4	251969	17 <td>CEY79H2</td> <td>Caenorhabditis elegans</td> <td>1.88e+01</td>	CEY79H2	Caenorhabditis elegans	1.88e+01

ALIGNMENTS

RESULT	1	G05576	307 bp	DNA	STS	04-JUN-1996
LOCUS		G05576				
DEFINITION		human STS WI-8386.				
ACCESSION		G05576				
KEYWORDS		STS sequence; primer; sequence tagged site.				
SOURCE		human STS derived from sequences in dbEST and the Unigene collection.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryotes: mitochondrial eukaryotes: Metazoa: Chordata: Vertebrates: Eutheria: Primates: Catarrhini: Homidae: Homo.				
AUTHORS		1 (bases 1 to 307)				
TITLE		Hudson, T.				
JOURNAL		Whitehead Institute/MIT Center for Genome Research: Physically				
COMMENT		Unpublished (1995)				

Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: TTTCCTTCTTAAACCTTCACC
Primer B: CCCTCTGCTGACACCTTGAT
STS size: 257
PCR Profile:
Presoak:
Denaturation:
Annealing: 56 degrees C

Feature	Human	Mouse	Score	DB	Length
Intron	1001..2309	1001..2309	32.2%	DB	10772
exon	/gene="Cp1"	/gene="Cp1"	Best Local Similarity 17.9%	Pred. No. 1.26e-01	
CDS	2310..2426	2310..2426	Matches 10; Conservative 32; Mismatches 14; Indels 0; Gaps 0;		
Intron	/number=2	/number=2			
exon	join(2328)..2426,6476..6690,6751..7462)	join(2328)..2426,6476..6690,6751..7462)			
misc.feature	/gene="Cp1"	/gene="Cp1"			
intron	/codon_start=1	/codon_start=1			
misc.feature	/product="cysteine proteinase-1"	/product="cysteine proteinase-1"			
intron	/db_xref="PID:g2305221"	/db_xref="PID:g2305221"			
misc.feature	/translation="MRTAVLLPLALLAAVAQVADYVMEEMHTFKLEHRKNTODET	/translation="MRTAVLLPLALLAAVAQVADYVMEEMHTFKLEHRKNTODET			
intron	EEERLRTFENENKHKIAKNHQFPAEKGVSFKAVNKYADLLIHEEROLMNGENTLH	EEERLRTFENENKHKIAKNHQFPAEKGVSFKAVNKYADLLIHEEROLMNGENTLH			
exon	OLFRASDSEFGVTEFISPAHYTLKSYDMRTKAVYIAVKQOGCGCMFSSIGALEGG	OLFRASDSEFGVTEFISPAHYTLKSYDMRTKAVYIAVKQOGCGCMFSSIGALEGG			
intron	HFRSGDVLVLSLEQNLVDCSTKYGNNCGMADNAFYRIKDNQIIDEKSPYEALIEG	HFRSGDVLVLSLEQNLVDCSTKYGNNCGMADNAFYRIKDNQIIDEKSPYEALIEG			
exon	EPCCDNKGTGATDRGFTDIPQDEKMKMAEAVATGVPVSAIDASHESFORYSGVYN	EPCCDNKGTGATDRGFTDIPQDEKMKMAEAVATGVPVSAIDASHESFORYSGVYN			
intron	ASSYPLV"	ASSYPLV"			
exon	2427..6475	2427..6475			
misc.feature	/gene="Cp1"	/gene="Cp1"			
intron	4546..4553	4546..4553			
misc.feature	/gene="Cp1"	/gene="Cp1"			
intron	/note="insertion site of P[CaSper](50C)"	/note="insertion site of P[CaSper](50C)"			
exon	6476..6690	6476..6690			
intron	/number=3	/number=3			
exon	6691..6750	6691..6750			
intron	/gene="Cp1"	/gene="Cp1"			
exon	6751..7707	6751..7707			
intron	/gene="Cp1"	/gene="Cp1"			
exon	/number=4	/number=4			
intron	join<8110..9300,9370..>9532)	join<8110..9300,9370..>9532)			
exon	/product="phenylalanyl tRNA synthetase"	/product="phenylalanyl tRNA synthetase"			
intron	join(8110..9300,9370..>9532)	join(8110..9300,9370..>9532)			
exon	/note="potential orf"	/note="potential orf"			
intron	/codon_start=1	/codon_start=1			
exon	/product="phenylalanyl tRNA synthetase"	/product="phenylalanyl tRNA synthetase"			
intron	/db_xref="PID:g2305222"	/db_xref="PID:g2305222"			
exon	/translation="MLTLRVQGARHVLKSTROCLASAPAKSPSSPOLLEYSGSTYA	/translation="MLTLRVQGARHVLKSTROCLASAPAKSPSSPOLLEYSGSTYA			
intron	TDGNTVPTKLLSYGVANGKHLQTDHPLESTIRQIVVFGAVGNORGNGLEFVYDGM	TDGNTVPTKLLSYGVANGKHLQTDHPLESTIRQIVVFGAVGNORGNGLEFVYDGM			
exon	EVVTVQONFENMLIPADHVSROKSDCYINOHLRAHTTAHVLEISGLDNFLVYGM	EVVTVQONFENMLIPADHVSROKSDCYINOHLRAHTTAHVLEISGLDNFLVYGM			
intron	LYRDEIDSTHYEPFHQADAVRLVTRDKLFEENPCLFEFEETSGTADPXLILPH	LYRDEIDSTHYEPFHQADAVRLVTRDKLFEENPCLFEFEETSGTADPXLILPH			
exon	YSWKPNPSPATRAVRAKLEMEHMKHYLXGLTKDLDFGRIRYRWVDFYFEPTOPSMLEI	YSWKPNPSPATRAVRAKLEMEHMKHYLXGLTKDLDFGRIRYRWVDFYFEPTOPSMLEI			
intron	YFKNMLLEVAGCGIMRHEITLQORSGVNDQSLQIYAGVLEIRLAVLFDIPDIRLFWSDS	YFKNMLLEVAGCGIMRHEITLQORSGVNDQSLQIYAGVLEIRLAVLFDIPDIRLFWSDS			
exon	GFELQSEKDLHNLPRKTPILSHYPOCTNDSFWLPQDIEVDAGSFNDPFDLVRSVAG	GFELQSEKDLHNLPRKTPILSHYPOCTNDSFWLPQDIEVDAGSFNDPFDLVRSVAG			
intron	DMVQIQLVDFKHPKTKGSSVCFRIYRMRMERTILQAEVNEITHQIASVSDSENVQ	DMVQIQLVDFKHPKTKGSSVCFRIYRMRMERTILQAEVNEITHQIASVSDSENVQ			
exon	IX"	IX"			

```

JOURNAL      Unpublished (1998)
REFERENCE    2 (bases 1 to 83675).
AUTHORS      Waterson,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (12-JUN-1998) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
              3 (bases 1 to 83675)
              Waterson,R.
COMMENT       Direct Submission
              Submitted (19-SEP-1998) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              On Sep 19, 1998 this sequence version replaced gi:3213167.
              SUBMITTED BY: WUGSC
              Genome Sequencing Center
              Department of Genetics
              Washington University
              St. Louis MO 63108, USA
              http://genome.wustl.edu/gsc
              mailto:saplens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GRB/CHR7, send
mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-1, prepared by
Plater de Jong and coworkers at Roswell Park Cancer Institute,
using the method described by Ioannou et al., Nature Genetics,
6:84-9 (1994). The library is from one male donor. For further
details, see http://bacpac.med.buffalo.edu/
The clone is available from Genome Systems, Inc.
(http://www.genomesystems.com).
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is DJ0724E13, 200 bp overlap.
Actual start of this clone is at base position 1 of DJ0537J23;
actual end is at 42321 of DJ0724E13.
Location/Qualifiers
1..83675
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="7"
  /map="7q11.2-p12"
  /clone_1lb="RPCI-1"
  /clone="DJ0537J23"
  1..276
    /rpt_family="L1"
repeat_region
    274..334
    /rpt_family="L1"
repeat_region
    335..1287
    /rpt_family="L1"
repeat_region
    1289..1313
    /rpt_family="L1"
repeat_region
    1315..2975
    /rpt_family="(TAA)n"

```


[illegible]

JOURNAL

Submitted (24-OCT-1997) to the DDBJ/EMBL/GenBank databases.
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
(E-mail: ynakamu@kazusa.or.jp, Tel: +81-438-52-3935,
Fax: +81-438-52-3934)

FEATURES

Location/Qualifiers
1..50641

/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MTG13"
/clone_11b="Mitsui P1"

BASE COUNT 16566 a 8254 c 8714 g 17107 t
ORIGIN

Query Match 31.0%; Score 27; DB 23; Length 50641;
Best Local Similarity 70.3%; Pred. No. 3.53e-01;
Matches 45; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Db 12705 TTTTAAAGTATATATATTAATTTGTAATTTTATTTATTAATATTATATATA 12764

OY 23 TTTAGAGGAGAAATATAGTAAATGTATTTGCTTAATAAATCATCTATACA 82

Db 12765 TATA 12768

OY 83 TCTA 86

RESULT 9
LOCUS HS11D6 128330 bp DNA HTG 08-OCT-1998
DEFINITION Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 11D6;

ACCESSION AL031056
NID 93724208

KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 128330)
Mashregui-Mohammadi, M.

JOURNAL Direct Submission
Submitted (08-OCT-1998) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humanyesanger.ac.uk Clone requests: clonerequestsanger.ac.uk

COMMENT On Oct 9, 1998 this sequence version replaced g1:3676169.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
segments. Unfinished: d0111D6 Contig_ID: 01780 acc=AL031056
Length: 128330 bp.

*** WARNING: Phase 1 High Throughput Genome Sequence ***

* This sequence is unfinished. When sequencing is complete,
* the sequence data presented in this record will be replaced
* by a single finished sequence with the same accession number.
Location/Qualifiers
1..128330
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="11D6"

BASE COUNT 37388 a 25652 c 26823 g 38467 t
ORIGIN

Query Match 31.0%; Score 27; DB 17; Length 128330;
Best Local Similarity 72.1%; Pred. No. 3.53e-01;
Matches 44; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Db 83292 AGATTATTTAAAGTAAATTCACATTTCACAAATCATTTCATTGTAATCTACGCA 83351
Cp 73 AGATTATTTATAGACAAATTTACACATTTCATTTCCTTTCTTAACACCTACCA 14

Db 83352 T 83352

Cp 13 T 13

RESULT 10
LOCUS AF012089 10772 bp DNA INV 05-AUG-1997
DEFINITION Drosophila melanogaster cysteine proteinase-1 (Cp1) gene, complete
cds, and phenylalanyl tRNA synthetase gene, partial cds.
ACCESSION AF012089
NID 92305220

KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Eukaryotes; Metazoa; Arthropoda;
Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 4546 to 4553)
Gray, Y.H.M., Tanaka, M.W. and Sved, J.A.
P-element-induced recombination in Drosophila melanogaster: hybrid
element insertion
Genetics 144 (4), 1601-1610 (1996)

JOURNAL MEDLINE
97132596
2 (bases 1 to 10772)
Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.
Structure of the cysteine proteinase (Cp1) gene of Drosophila
melanogaster and associated mutational effects
Unpublished

REFERENCE 3 (bases 1 to 10772)
Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.
Direct Submission
Submitted (30-JUN-1997) School of Biological Sciences, University
of Sydney, Biology A12, Sydney University, NSW 2006, Australia

JOURNAL Location/Qualifiers
1..10772
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
Join(872..1000,2310..2426,6476..6690,6751..7707)
/gene="Cp1"
/product="cysteine protease"

gene 872..7707
/gene="Cp1"
872..1000
/gene="Cp1"

exon /number=1
1001..2309
/gene="Cp1"
2310..2426
/gene="Cp1"

intron /number=2
Join(2328..2426,6476..6690,6751..7462)
/gene="Cp1"

CDS /codon_start=1
/product="cysteine proteinase-1"
/db_xref="PID:g2305221"

translation="MRTAVLLPLIALVAQAVSFADVVEWHTEFLERKNYODET
EERFLKIFENKRIKAKHNOFAEGKVSFLAVKADLLHFEFLMNGFNTLTK
QLRADESPKGVTFISPAHYTLPRSVPMRTGAVTAVKDGSGSCWAFSTGLDGO
HFRKSGVLVSEQNLVDCSTKYGNNCGNDMNAFRYIKDNGSIDTEKSYPEALD
DSCHFNKGTGATDRGFTDIPQDEKMAAATVGVSAVIDASHSFPTSEGVN
EPCDDAONLDHGVLVYFGTDESEEDYLVKNSGITWGDKFTIKLRNECCGLAS
ASSIPLY"

misc_feature
intron
2427..6475
/gene="Cp1"
4546..4553
/gene="Cp1"
/note="insertion site of p(Casper)(50C)"
/citation=[1]

	rRNA	/mitochondrion /db_xref="taxon:4932"
	CDS	1..411 "/note="tRNA 155" 787..1026 "/note="RF2; unidentified reading frame; Protein sequence is in conflict with the conceptual translation" /codon_start=1 /transl_table=3 /db_xref="PIR:E5941" /db_xref="PID:g133448" /translation="IMNYKKXGNNWFLNMFNKATNMVTELLLYNNYINMFKIKT RPDSEGVKPAAGACGTILYLINDKSIIINKRIYN" 820..822 "/note="translated tga codon (tp)" 1726..1793 "/note="tRNA-Trp"
BASE COUNT	1058 a 246 c 241 g 1061 t	
ORIGIN		
Query Match	28.7%; Score 25; DB 23; Length 2606;	
Best Local Similarity	70.0%; Pred. No. 2,65e+00;	
Matches	42; Conservative 0; Mismatches 18; Indels 0; Gaps 0;	
Df	538 GAAGCAGAATAATTAAATTAATTTAATATATAATAATAATAATAATA 597	
Oy	28 GAAGCAAAAATAATAGTAAAACTGGTAATTTCCTTAATAAATCATCTACTAACATCTAA 87	
RESULT 14	MISC37 3980 bp DNA PLN 31-MAR-1992	
LOCUS	X00845	
DEFINITION	yeast mitochondrial genes for 15S rRNA and tRNA-Trp.	
ACCSSION	g13557	
NID	15S ribosomal RNA; ribosomal RNA; transfer RNA; transfer RNA-Trp; unidentified reading frame. baker's yeast.	
SOURCE	Mitochondrion Saccharomyces cerevisiae	
ORGANISM	Eukaryota; mitochondria eukaryotes; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces. 1 (bases 1 to 2606) Martin,R.P., Siber,A.P., Bordome,R. and Dirheimer,G. Yeast mitochondrial tRNAs: structure, coding properties, and gene organization Mol. Biol.(Mosk.) 17, 915-936 (1983)	
REFERENCE	MoI. Biol.(Mosk.) 17, 915-936 (1983)	
AUTHORS	L.I.M., Tzagoloft,A.; Underbrink-Lyon,K. and Martin,N.C.	
TITLE	Identification of the paromycin-resistance mutation in the 15 S rRNA gene of yeast mitochondria J. Biol. Chem. 257 (10), 5921-5928 (1982)	
JOURNAL MEDLINE	82167455	
FEATURES	Location/Dualifiers	
Source	1..3980 /organism="Saccharomyces cerevisiae" /mitochondrion /db_xref="taxon:4932"	
BASE COUNT	1655 a 368 c 425 g 1532 t	
ORIGIN		
Query Match	28.7%; Score 25; DB 23; Length 3980;	
Best Local Similarity	70.0%; Pred. No. 2,65e+00;	
Matches	42; Conservative 0; Mismatches 18; Indels 0; Gaps 0;	
Df	1913 GAAGCAGAATAATTAAATTAATTTAATATATAATAATAATAATAATA 1972	
Oy	28 GAAGCAAAAATAATAGTAAAACTGGTAATTTCCTTAATAAATCATCTACTAACATCTAA 87	
RESULT 15	YSGMTGCI1 4350 bp DNA PLN 26-MAY-1995	
LOCUS	YC00845	
DEFINITION	saccharomyces cerevisiae mitochondrion 15S ribosomal RNA (15S rRNA) gene; ORF3; Transfer RNA-Trp (tRNA-Trp) gene.	

```

ACCESSION      L36895 M62622
NID            9559271
KEYWORDS
SEGMENT
SOURCE
ORGANISM
REFERENCE
  1 (bases 1 to 4350)
  de Zamaroczy,M. and Bernardi,G.
  The primary structure of the mitochondrial genome of Saccharomyces
  cerevisiae--a review
  Gene 47 (2-3), 155-177 (1986)
  87163488
FEATURES
  source
    1..4350
    /organism="Saccharomyces cerevisiae"
    /mitochondrion
    /db_xref="taxon:4932"
    486..2155
    /gene="15S rRNA"
    /product="15S ribosomal RNA"
    486..2155
    /gene="15S rRNA"
    2531..2773
    /note="ORF3"
    /codon_start=1
    /transl_table=3
    /db_xref="PID:9567916"
    /translation="MMNYKYEKGNWELNMFKNYNNMYTLLLYNNYINMFNRIKTTI
    RPDSFGVPAGAACTLPILLYINDKSLIINKFIYNI"
    3467..3537
    /gene="tRNA-Trp"
    /product="transfer RNA-Trp"
    3467..3537
    /gene="tRNA-Trp"
  BASE COUNT      1771 a      420 c      477 g      1682 t
  ORIGIN
    28 7%: Score 25; DB 23; Length 4350;
    Best Local Similarity 70.0%; Pred. No. 2,65e+00;
    Matches 42; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
  Db 2282 GAAAGAGAAAATATAAATATAATATATATATATATAATAATAATATAA 2341
    ||||| ||||| | | |||| | ||||| || | | || |||
  Oy 28 GAAAGAAAATATAGTAAATGTGTATTTGTCTTATAAATCTATCTTACATCTAA 87

```


FEATURES	source	Osaka University 3-1, Yamadaoka Suita, Osaka, 565 Japan Phone: 06-877-5111 Fax : 06-875-1922. Location/Qualifiers 1. .62
BASE COUNT	30 a 11 c 9 g 11 t 1 others	
ORIGIN		
Query Match	100.0%; Score 61; DB 7; Length 62;	
Best Local Similarity	100.0%; Pred. No. 3,15e-29;	
Matches	62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Db	1 GATCCCATCTTTTCAAAACAAATATAAACAATAAGACTGCAGAGACTGANGGA 60	
Qy	1 GATCCCATCTTTTCAAAACAAATATAAACAATAAGACTGCAGAGACTGANGGA 60	
Db	61 AA 62	
Qy	61 AA 62	
RESULT	2 HUMGS01613 58 bp mRNA EST 17-JUN-1996	
LOCUS	Human HL60 3'directed MD01 cDNA, HUMGS01613, clone pm1209, mRNA	
DEFINITION	sequence.	
ACCESSION	D20638	
NCBI	9501734	
KEYWORDS	EST; EST(expressed sequence tag); HL60; gene signature(GS); granulocyte; macrophage; monocyte; promyelocyte.	
SOURCE	Homio sapiens adult female promyelocyte cell_line:HL60 cDNA to mRNA	
ORGANISM	Homio sapiens	
REFERENCE	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homio.	
REFERENCE	1 (sites)	
AUTHORS	Okubo,K., Itoh,K., Fukushima,A., Yoshii,J. and Matsubara,K.	
TITLE	Monitoring cell physiology by expression profiles and discovering cell type-specific genes by compiled expression profiles	
JOURNAL	Genomics 30 (2), 178-186 (1995)	
MEDLINE	96163870	
REFERENCE	2 (bases 1 to 58)	
AUTHORS	Okubo,K., Fukushima,A., Yoshii,J., Niijima,T., Kojima,Y., Yoshinari,H., Arimoto,O. and Matsubara,K.	
TITLE	Gene expression of human promyelocytic cell line HL60before and after induction of differentiation. A new application of 3'directedcDNA sequencing	
JOURNAL	Unpublished (1993)	
REFERENCE	3 (bases 1 to 58)	
AUTHORS	Okubo,K.	
TITLE	Direct Submission	
JOURNAL	Submitted (10-Sep-1993) to the DDBJ/EMBL/Genbank databases. Kousaku Okubo, Osaka University, Institute for Molecular and Cellular Bio: 1-3,Yamada-oka, Suita, Osaka pref. 565, Japan (E-mail:kousaku@imcb.osaka-u.ac.jp, Tel:06-877-5111(ex.3315), Fax:06-877-1922)	
COMMENT	Submitted (10-Sep-1993) to DDBJ by: Kousaku Okubo Institute for Molecular and Cellular Biology Osaka University 3-1, Yamadaoka Suita, Osaka, 565 Japan Phone: 06-877-5111 Fax : 06-875-1922.	

```

FEATURES
SOURCE
    Location/Qualifiers
        1..58
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_line="HL60"
            /cell_type="promyelocyte"
            /dev_stage="adult"
            /sex="female"

BASE COUNT      30 a       9 c       8 g       9 t       2 others
ORIGIN
Query Match          62.3%; Score 38; DB 7; Length 58;
Best Local Similarity 91.8%; Pred. No. 5.8e-11;
Matches 45; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Db      11 TTTTAAACCAATTAACAATAAAGCTGCAAG-AAAGCTGAGNAA 58
|||||
|||
14 TTCAAAAACAAATTAACATAAAGCTGCAAGAAGAACTGANGANA 62

RESULT      3
LOCUS      A1185894      443 bp      mRNA      EST      09-OCT-1998
DEFINITION ge45c11.x1 Soares_fetal_lung_NBHL19W Homo sapiens CDNA clone
IMAGE:1141940 3', mRNA sequence.
ACCESSION  A1185894
NID        93736532
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 443)
AUTHORS   NC1-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL   Tumor Gene Index
COMMENT   Unpublished (1997)

FEATURES
SOURCE
    Contact: Robert Strausberg, Ph.D.
    Tel:(301) 496-1550
    Email: Robert.Strausberg@nih.gov
    This clone is available royalty-free through LNL ; contact the
    IMAGE Consortium (info@image.lnl.gov) for further information.
    Seq primer: -40UP from Gibco
    High quality sequence stop: 445.
    Location/Qualifiers
        1..443
            /organism="Homo sapiens"
            /note="Organ: Lung; Vector: pRT73D (Pharmacia) with a
            modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
            strand cDNA was primed with a Not I - oligo(dT) primer
            [5'-TGTTACCACATCTGATGGAGGAGCGGCCGCATTTTGTGTGTGT-3']
            double-stranded cDNA was size selected, ligated to Eco RI
            adapters (Pharmacia), digested with Not I and cloned into
            the Not I and Eco RI sites of a modified pRT73 vector
            (Pharmacia). Library went through one round of
            normalization to a Cot = 5. Library constructed by Bento
            Soares and M Fatima Bonaldo. This library was constructed
            from the same fetus as the fetal heart library, Soares
            fetal heart NBHL19W."
            /db_xref="taxon:9606"
            /clone_id="IMAGE:1141940"
            /clone_lib="Soares_fetal_lung_NBHL19W"
            /dev_stage="19 weeks"
            /lab_host="DH10B (ampicillin resistant)"

BASE COUNT      163 a      73 c      106 g      101 t
ORIGIN
Query Match          44.3%; Score 27; DB 18; Length 443;
Best Local Similarity 78.3%; Pred. No. 3.72e-03;
Matches 36; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db      171 AAAAACCAAAACAAAAAAGCAGCAGCAAGAAAAAGCGA 216
|||||
|||||

```


ACCESSION AA844688
 NID 92931139
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 407)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNML at:
 www.bio.lnml.gov/btrp/image/image.html
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 350.
 Location/Qualifiers
 1..407
 /organism="Homo sapiens"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech
 Laboratories, Inc., and primed with a Not I - oligo(dT)
 primer [5']
 TGTACCAATCTGAAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3']
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 /db_xref="taxon:9606"
 /clone="IMAGE:1376153"
 /sex="male"
 /sex="male"
 /lab_host="DH10B"
 /lab_host="DH10B"
 BASE COUNT 150 a 57 c 51 g 149 t
 ORIGIN
 Query Match 41.0%; Score 25; DB 13; Length 407;
 Best Local Similarity 93.1%; Pred. No. 7,46e-02;
 Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Db 130 TCTTTAGAAAACAAATTAACATTA 158
 OY 10 TCTTTCAAAACAAATTAACATTA 38
 RESULT 8
 LOCUS AA85450 246 bp mRNA EST
 DEFINITION vw73b01.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
 1260553 5', mRNA sequence.
 ACCESSION AA85450
 NID 92942988
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 246)
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

TITLE Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 JOURNAL Waterston, R.
 COMMENT The WashU-HMI Mouse EST Project
 Unpublished (1996)
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through ILNML; contact the
 IMAGE Consortium (info@image.lnml.gov) for further information.
 MGI:663105
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 230.
 Location/Qualifiers
 1..246
 /organism="Mus musculus"
 /strain="NIH/Swiss"
 /note="Organ: heart; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dr. 93 pooled NIH/Swiss 13 day embryo hearts.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
 adaptor sequence: 5' CAATTCGACGAG 3' -3' adaptor
 sequence: 5' CTCGACTTCTTTTCTTTTCTTTT 3' "
 /db_xref="taxon:10090"
 /clone="1260553"
 /clone_lib="Stratagene mouse heart (#937316)"
 /sex="pooled"
 /tissue_type="heart"
 /dev_stage="13 day embryos"
 /lab_host="SOLR (kanamycin resistant)"
 BASE COUNT 106 a 52 c 50 g 38 t
 ORIGIN
 Query Match 39.3%; Score 24; DB 13; Length 246;
 Best Local Similarity 74.5%; Pred. No. 3.20e-01;
 Matches 35; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
 Db 6 CAAATAATTAATTAATTAAGAAAGAAAGAAAGAA 52
 OY 16 CAAATAATTAATTAATTAAGAAAGAAAGAAAGAA 62
 RESULT 9
 LOCUS H88078 359 bp mRNA EST
 DEFINITION yw20d10.r1 Homo sapiens cDNA clone 252787 5'.
 ACCESSION H88078
 NID 91069657
 KEYWORDS EST.
 SOURCE human clone-252787 primer-M13R1 library-Morton Fetal Cochlea
 vector-pBluescript SK-host-SOLR cells (kanamycin resistant).
 Rsite1-EcoRI Rsite2-XhoI The cDNA was oligo (dT) primed with an
 XhoI restriction enzyme recognition site and an 18 base poly dT
 sequence. For the 5' end, the synthesized cDNA termini were
 treated with T4 DNA polymerase and EcoRI adaptors were ligated to
 the blunt ends. adaptor linker: GAATTCGCCGACGAG.
 Homo sapiens
 Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata;
 Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
 Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
 Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 359)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisan, E.,
 Waterston, R., Williamson, A., Woldmann, P. and Wilson, R.
 WashU-Merck EST Project
 Unpublished (1995)
 TITLE JOURNAL
 COMMENT Contact: Wilson RK

CP 61 TTCCTCAGTCTTCTGAGCTTATTTGTTTATTTGTTTGA 15

RESULT 12
LOCUS AA814930 381 bp mRNA EST
DEFINITION OC07C03.s1 NCI_CGAP_GCB1 Homo sapiens CDNA clone IMAGE:1340164 3',
RNA sequence:
AA814930
NID g2884526
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote: Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 381)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbfp/image/image.html
Insert Length: 633 Std Error: 0.00
Seq primer: -40m3 fwd. ET from Amersham
High quality sequence stop: 370.
Location/Qualifiers
1. 381

FEATURES
source
/organism="Homo sapiens"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, 19D-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). CDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTACCAATCTGACAGCGAGCGCGCCCTCATTTTTTTTTTTTTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. Library
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_GCB1"
/clone_id="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
BASE COUNT 110 a 77 c 46 g 148 t
ORIGIN

Query Match 39.3%; Score 24; DB 13; Length 381;
Best Local Similarity 74.5%; Pred. No. 3.20e-01;
Matches 35; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Db 334 TTCCACAGGATTCCTTAACTTTTGTGTTTGTGTTTGA 380
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 61 TTCCCTCAGTCTTCTGAGCTTATTTGTTTATTTGTTTGA 15

RESULT 13
LOCUS AA401809 403 bp mRNA EST
DEFINITION z65511.s1 Scores total fetus NB2HF8 9w Homo sapiens CDNA clone
758564 3', mRNA sequence.

ACCESSION AA401809
NID g2055845
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote: mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae;
Homo.
REFERENCE 1 (bases 1 to 403)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,
Moore, B., Schellenberg, K., Stepec, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-Merck EST Project 1997
Unpublished (1997)
COMMENT

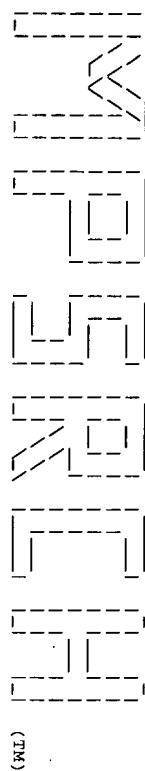
CONTACT: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 322.
Location/Qualifiers
1. 403

FEATURES
source
/organism="Homo sapiens"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'-
TGTACCAATCTGACAGCGAGCGCGCCCTCATTTTTTTTTTTTTTTT-3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. Library
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_lib="Soares total fetus NB2HF8 9w"
/clone_id="758564"
/dev_host="8-9 weeks"
/lab_host="DH10B"
complement(<1..>403)
BASE COUNT 70 a 84 c 74 g 175 t
ORIGIN

Query Match 39.3%; Score 24; DB 6; Length 403;
Best Local Similarity 85.3%; Pred. No. 3.20e-01;
Matches 29; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 370 GTGTTTGTGTTTGTGTTTGTGTTTAAAAAAT 403
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 41 GTCTTATGTTTATTTGTTTGTGTTTGAAGAAAT 8

RESULT 14
LOCUS AA774222 408 bp mRNA EST
DEFINITION ab55f11.s1 Stragene lung carcinoma 937218 Homo sapiens CDNA clone
844749 3', similar to contains Alu repetitive element; contains
element MRR26 repetitive element; mRNA sequence.
ACCESSION AA774222
NID g2825520
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote: Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 408)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kilzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_n n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Tue Dec 22 05:03:59 1998; MasPar time 27.06 Seconds
311.736 Million cell updates/sec
Tabular output not generated.

Title: >US-08-530-112A-1449
Description: (1-62) from US08530112A.seq
Perfect Score: 61
N.A. Sequence: 1 GATCCCAATCTTTTCAGAA.....GCAAGAGACTGAGGAAA 62
Comp: CTAGGGGTAAGAAAGTTT.....CGTCCCTCTGACTNCTTT

Scoring table:
Gap 6
TABLE default

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 6.582; Variance 6.199; scale 1.062

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description	Pred. No.
1	61	100.0	62.19	T20449	Human gene signature
2	35	57.4	91.9	Q51746	Oligonucleotide probe
3	25	41.0	68.32	T63355	Messenger RNA primer
4	25	41.0	68.34	T73397	Oligonucleotide tag c
5	25	41.0	70.24	T14325	Conjugate formed by 1
6	24	39.3	204.1	N81164	Base substituted Eco
7	24	39.3	1030.9	O55113	Streptococcal fibrino
8	23	37.7	830.33	T62455	Lambda 5h-1 clone con
9	23	37.7	4100.16	Q74062	The rat beta-actin ge
10	22	36.1	204.1	N81164	Base substituted E.co
11	22	36.1	251.19	T20099	Human gene signature
12	22	36.1	1689.3	Q22205	Murine mast cell grow
13	22	36.1	7997.15	Q89553	Rat cholesterol 7 alp

14	22	36.1	7997.34	T79737	Rat cholesterol 7- alp
15	21	34.4	427.35	T65071	Canine genomic micros
16	21	34.4	427.35	T65068	Canine genomic micros
17	21	34.4	789.7	O44255	Partial bovine VEGF-1
18	21	34.4	789.2	O10792	Bovine vascular endot
19	21	34.4	829.2	O10796	Bovine vascular endot
20	21	34.4	961.2	O10791	Bovine vascular endot
21	21	34.4	961.7	O44259	Bovine VEGF-164 codin
22	21	34.4	2460.30	T67106	Murine lymphoid enhan
23	21	34.4	3129.31	T68667	Rat growth hormone se
24	21	34.4	3129.31	T69759	Rat growth hormone se
25	21	34.4	3850.3	O11093	Chitinase gene contai
26	21	34.4	3850.3	O15147	Basic chitinase gene
27	21	34.4	9278.13	O76124	Human MDC genomic DNA
28	21	34.4	10965.4	O28507	Tomato vacuolar inver
29	21	34.4	40352.40	V02032	MAGE-B cluster DNA se
30	20	32.8	64.11	O68613	TENAPOLYU primer.
31	20	32.8	163.19	T20243	Human gene signature
32	20	32.8	300.8	O59905	Human brain Expressed
33	20	32.8	347.8	O60393	Human brain Expressed
34	20	32.8	880.7	O43675	Sequence of Intron 25
35	20	32.8	1374.40	V02068	Human nlnjurin 1 vari
36	20	32.8	1442.15	O90211	Brassica condenseng e
37	20	32.8	1450.3	O20843	Murine mast cell grow
38	20	32.8	1568.35	T94534	Alfalfa cinamoyl COA
39	20	32.8	1611.11	O69159	Human Fc-alpha-R cDNA
40	20	32.8	1728.22	T35868	Prostate-specific ant
41	20	32.8	2272.20	T11104	Transforming growth f
42	20	32.8	4287.40	V20475	Human AML1/MTG8 oncog
43	20	32.8	4287.11	O58996	AML1-MTG8 fusion.
44	20	32.8	5921.7	O43701	Sequence of the human
45	20	32.8	14356.15	O90511	CEA genomic clone.

ALIGNMENTS

RESULT 1
ID T20449 standard; cDNA to mRNA; 62 BP.
AC T20449;
AT 19-JUL-1996 (first entry)
DE Human gene signature HUMGS01603.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
OS Homo sapiens.
PN MO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1995; JP-355504.
PA (MATSU) MATSUBARA K.
PI (OKUBO) OKUBO K.
PI Matsubara K, Okubo K;
DR MPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1: Page 639; 2245pp; Japanese.
CC A single-stranded cDNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridize to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridize with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.

```

SQ      Sequence      62 BP;      30 A;      11 C;      9 G;      11 T;

Query Match
Best Local Similarity 100.0%; Score 61; DB 19; Length 62;
Matches      62; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Db      1 gatcccatcttccaaaacaataaacaataaagactgtcgaaggaaactcganga 60
      |||
Oy      1 GATCCCATCTTTTTCAAAAACAATAAACAATAAAGACTGCAAGAACTGANGGA 60

Db      61 aa 62
      ||
Oy      61 AA 62

RESULT      2
ID      Q51746 standard; cDNA; 91 BP.
AC      Q51746;
DT      31-MAY-1994 (first entry)
DE      Oligonucleotide probe MK14-A
KM      Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW      ss.
OS      Synthetic.
PN      EP-571911-A.
PD      01-DEC-1993.
PF      24-MAY-1993; 108325.
PR      26-MAY-1992; US-889651.
PA      (BECT ) BECTON DICKINSON CO.
DR      Shank DD, Spears PA;
PT      New oligo:nucleotide probes specific for Mycobacteria - used for
PT      detection and amplification of Mycobacteria nucleic acid in
PT      samples
PS      Claim 3; Page 14; 23pp; English.
CC      Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC      (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC      cross reacted to a few non-mycobacterial spp. The probe may
CC      be useful as an initial screen for mycobacterial infection.
CC      See also Q51735-45 and Q51747-59.
CO      Sequence      91 BP;      5 A;      17 C;      15 G;      4 T;

Query Match
Best Local Similarity 0.0%; Score 35; DB 9; Length 91;
Matches      0; Conservative      42; Mismatches      7; Indels      0; Gaps      0;

Db      12 svhsyyvvhvshhsyvhvvhvhsvvvvhvvhvvhvhyvysv 60
      :
Oy      1 GATCCCATCTTTTCAAAAACAATAAACAATAAAGACTGCAAGGA 49

RESULT      3
ID      T63253 standard; DNA; 68 BP.
AC      T63253;
DT      07-OCT-1997 (first entry)
DE      Messenger RNA primer containing a complement tag sequence.
KM      mRNA: oligonucleotide tag; hybridisation; automated DNA mapping;
OS      genetic identification; polynucleotide; target; ss.
OT      Synthetic.
FH      key
FT      modified_base      1
FT      location/Qualifiers
FT      /*tag= a
FT      /note= "5'-labelled with biotin"

W09641011-A1.
PD      19-DEC-1996.
PF      06-JUN-1996; U09513.
PR      12-OCT-1995; WO-U12791.
PR      07-JUN-1995; US-478238.
PA      (SPEC-) SPECTRAGEN INC.
PI      Albrecht G, Brenner S;
PT      WPI; 97-099943/09.
PT      Sorting poly:nucleotide(s) then to solid supports by attachment to
PT      oligo:nucleotide tags - on specific hybridisation of tags to
PT      immobilised complement, e.g. for automated DNA mapping and

```

```
PT sequencing, genetic identification and diagnosis
PS Disclosure: Page 22; 79pp; English.
CC A method of sorting a population of polynucleotides on to one or more
CC solid supports has been produced. The polynucleotides are sorted on to
CC solid supports by: (a) attaching an oligonucleotide tag to each
CC polynucleotide, each tag being minimally cross-hybridising; (b) sampling
CC the population so that all different polynucleotides have different
CC tags; (c) sorting by specific hybridisation of the tags with their
CC complements which are attached as uniform populations of identical
CC oligonucleotides in spatially distinct regions on one or more solid
CC phase support. The present sequence represents a primer, from a set,
CC preferably containing complements of tag sequences. When the population
CC of polynucleotides consists of messenger RNA, oligonucleotide tags may
CC be attached by reverse transcribing the RNA with the set of primers
CC including the present sequence. The method can be used to identify a
CC population of mRNA molecules; to detect presence/absence of selected
CC target sequences in a target polynucleotide; or to identify
CC polynucleotides, including new ones in cDNA libraries, e.g. for
CC construction and use of combinatorial chemical libraries; large scale
CC DNA mapping and sequencing; genetic identification; medical diagnosis
CC (e.g. analysis of gene expression in diseased and normal tissue). The
CC method is easily automated for manipulation and sorting of
CC polynucleotides in large scale parallel processing where many target
CC polynucleotides, or many target segments of a single polynucleotide, are
CC sequenced simultaneously.
SQ Sequence 68 BP; 3 A; 13 C; 4 G; 21 T;

Query Match          41.0%; Score 25; DB 32; Length 68;
Best Local Similarity 43.5%; Pred. No. 6.81e-01;
Matches    20; Conservative   15; Mismatches 11; InDels    0; Gaps    0;

Db      21 www.cmcwww.cmcwww.cmcwww.cmcwwwggtttttttttttttttt 66
Cp      62 TTTCCTCAGTCTCCTGCAGCTTTATGTATTGTTTT 17

RESULT      4
ID      T73397 standard; DNA; 68 BP.
AC      T73397;
DC      03-DEC-1997 (first entry)
DE      Oligonucleotide tag containing primer #1 for toxicity determination.
KW      PCR; primer: amplify; polymerase chain reaction; toxicity determination;
KM      oligonucleotide tag; hepatocyte; mRNA populations fingerprinting; ss.
OS      Synthetic.
FH      key Location/Qualifiers
FI      modified_base I
FT      /*tag= a
FT      /note= "Biotin labelled"
FT      misc_feature 12..47
FT      /*tag= b
FT      /note= "Oligonucleotide tag"

NN09713877-AI.
PD      17-APR-1997.
PF      11-OCT-1996; U16342.
PR      06-JUN-1996; WO-U09513.
PR      12-OCT-1995; WO-U12791.
PA      (LYNX-) LYNX THERAPEUTICS INC.
PI      Martin DW;
DR      WPI: 97-23591/21.
PT      Massively parallel signature sequencing - useful to test toxicity of
PT compound, or to identify genes which are differentially expressed in
PT selected tissue or a test animal after treatment with a compound
PS disclosure; Page 26; 65pp; English.
CC T73397-T73400 represent amplification primers that can be used in the
CC method of the invention. The method is for determining the toxicity of a
CC compound. The method comprises administering a compound to a test
CC organism, extracting mRNA molecules from 1 or more tissues, and forming a
CC population of cDNA molecules from each mRNA population. Each cDNA
CC molecule has attached an oligonucleotide tag, and then each population of
CC cDNA molecules is then sampled. The cDNA molecules are then sorted by
CC specifically hybridising the tags with their complements, which are
CC attached in spatially discrete regions on 1 or more solid phase supports.
CC The sequence of each of the sorted cDNA molecules is determined to form a
```

```

CC frequency distribution of expressed genes for each tissue. The frequency
CC of distribution of expressed genes is then correlated with the toxicity
CC of the compound. The method, which comprises the massively parallel
CC signature sequencing technique, can be used to test the toxicity of a
CC compound, or to identify genes which are differentially expressed in a
CC selected tissue of a test animal after treatment with a compound, in a
CC mammalian tissue culture, preferably comprising hepatocytes. The method
CC may also be used to fingerprint mRNA populations, either in isolated
CC measurements or in the context of dynamically changing population,
CC partial sequence information is obtained from a large sample, e.g. 10 to
CC 100000, or more, of cDNA attached to separate microparticles.
SQ Sequence 68 BP; 3 A; 13 C; 4 G; 21 T;

Query Match
Best Local Similarity 43.5%; Score 25; DB 34; Length 68;
Matches 20; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

DB 21 WWWWWWGAGTCTTCTGAGCTTTATTTATTTGTTTTT 66
CP 62 TTTCTTCACTCTTCTGAGCTTTATTTATTTGTTTTT 17

RESULT 5
ID T14325 standard; DNA; 70 BP.
AC T14325;
DE 16-JAN-1997 (first entry)
KW Conjugate formed by labelling and sorting of nucleotide molecules.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 65
FT /tag= a
FT /note= "This N represents the amplified cDNA
FT sequence."
FT
FT W06912039-A1.
PD 25-APR-1996.
PF 12-OCT-1995; U12678.
PR 13-OCT-1994; US-332348.
PA 19-DEC-1994; US-359295.
PA (LYNX-) LYNX THERAPEUTICS INC.
PI Brenner S;
DR WPI: 96-222023/22.
PT Labelling and sorting mols. using oligo:nucleotide tags - useful in
PT large-scale parallel operations; e.g. DNA sequencing and mRNA
PT fingerprinting
PS Disclosure: Page 20; 71pp; English.
CC Determining the nucleotide sequence (1) of a target polynucleotide
CC (T) comprises: (a) generating from T a plurality of fragments that
CC cover T; (b) attaching an oligonucleotide tag from a repertoire of
CC tags, to each fragment such that all the same fragments have the
CC same tag, and all different fragments have different tags; (c)
CC sorting the fragments by specifically hybridising the tags with
CC their respective tag complements; (d) determining (1) of a portion
CC of each of the fragments; and (e) determining (1) of T by collating
CC the sequences of the fragments. The tagging system can be used with
CC single base sequencing methods to sequence polynucleotides up to
CC several kilobases in length. The tagging system permits many
CC thousands of fragments of a target polynucleotide to be sorted onto
CC one or more solid phase supports and sequenced simultaneously.
CC A primer which is initially used to reverse transcribe mRNA is
CC described in T14322. If the amplified product is then needed to be
CC attached to a solid phase support a sequence like the one given in
CC T14323 may be used. The mRNA would then be removed and the second
CC strand of cDNA produced using a primer with a similar form to that
CC described in T14324. After restriction enzyme digestion, the
CC conjugate would have a formula similar to this.
SQ Sequence 70 BP; 2 A; 12 C; 3 G; 19 T;

Query Match
Best Local Similarity 41.0%; Score 25; DB 24; Length 70;
Matches 20; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

DB 17 WWWWWWGAGTCTTCTGAGCTTTATTTATTTGTTTTT 62

```

```

CP 62 TTTCTTCACTCTTCTGAGCTTTATTTATTTGTTTTT 17

RESULT 6
ID N81164 standard; DNA; 204 BP.
AC N81164;
DE 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT /tag= a
FT /function= multiple cloning site
FT primer_bind 187..204
FT /tag= b
FT
FT EP-285123-A.
PD 05-MAY-1988.
PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivu A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure: P; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match
Best Local Similarity 39.3%; Score 24; DB 1; Length 204;
Matches 1; Conservative 32; Mismatches 20; Indels 0; Gaps 0;

DB 131 bhddvvybbyvnnhnnccbnhvnbnhnnrnyvrrdrrd 183
CP 55 CAGTCTTCTTCTGAGCTTTATTTATTTGTTTTTGAAGAAGAGGGA 3

RESULT 7
ID 055113 standard; DNA; 1030 BP.
AC 055113;
DE 12-JUL-1994 (first entry)
DE Streptococcal fibrinogen binding protein gene.
KW Binding; fibronectin; Streptococcus; vaccine; antigen; probe; ss.
OS Streptococcus pyogenes.
FH Key Location/Qualifiers
FT cds 1..807
FT /tag= a
FT /product= Fibronectin binding protein
FT repeat_unit 971..980
FT /tag= b
FT /rpl_type= INVERTED
FT repeat_unit 986..995
FT /tag= c
FT /rpl_type= INVERTED
FT
FT W09401465-A.
PD 20-JAN-1994.
PF 13-JUL-1993; SE0625.
PR 14-JUL-1992; SE-002164.
PA (STRO-) STROEMQUIST MITJOEMEDICINSKONSULT AB.

```

[illegible]


```

RESULT 11
ID T20099 standard: cDNA to mRNA: 251 BP.
AC T20099:
DT 26-JUL-1996 (first entry)
DE Human gene signature HUMGS01242.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
human: cloning; mapping; non-biased library; diagnosis; detection;
cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PE 11-NOV-1994: J01916.
PR 12-NOV-1993: JP-355504.
PA (MATSU) MATSUBARA K.
PA (OKUBO) OKUBO K.
PA Matsubara K, Okubo K;
DR WPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1: Page 560; 2245pp. Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19901-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
CC Sequence 251 BP; 67 A; 27 C; 47 G; 107 T;

Query Match 36.1%; Score 22; DB 19; Length 251;
Best Local Similarity 92.3%; Pred. No. 7.94e+00;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0.

Db 19 ttttggttgtttgtttgaa 44
    ||| ||||| ||||| |||||
Cp 38 tttatgtttattttgtttgaaa 13

RESULT 12
ID Q22205 standard: cDNA; 1689 BP.
AC Q22205:
DT 30-APR-1992 (first entry)
DE Murine mast cell growth factor cDNA clone mMGF-10'.
KW Hematopoietin; interleukin; IL-3; c-kit oncogene; mouse;
KW proliferation; ss.
OS Mus musculus.
PN W09200376-A.
PD 09-JAN-1992.
PE 14-JUN-1991: U04274.
PR 25-JUN-1990: US-543264.
PR 10-AUG-1990: US-565840.
PR 28-AUG-1990: US-574152.
PR 21-SEP-1990: US-586073.
PR 12-JUN-1991: US-713715.
PA (IMMUNEX) IMMUNEX CORP.
PA Williams DE, Lyman S;
PI WPI: 92-04158/05.
PT New isolated DNA encoding human mast cell growth factor - useful in
PT stimulating proliferation and haematopoietic cells with growth factor,
PT to treat haemolytic and hypoproliferative anaemias
PS Disclosure: Page 37, 59pp; English.
CC This cDNA sequence is designated as SEQ. ID. NO. 2 in the specification
CC and is said to correspond to cDNA clone mMGF-10 given in Figure 2

```

Query Match	Best Local Similarity	89.3%	Pred. No. 7.94e+00;	Length 1689;
Matches	25; Conservative	0;	Mismatches	3; Indels 0; Gaps 0.
CC of the specification (see also Q20843). The two sequences, however, are not the same. See also Q20842-5 and Q22204-7.				
CC Sequence 1689 BP; 527 A; 338 C; 388 G; 436 T.				
SO				
Db 1599 tttaaaaaaacaaacaaacataaag 1626				
OY 12 TTTTCAAAAACAAATATAAACATAAAG 39				
RESULT 13				
ID Q89553 standard; DNA; 7997 BP.				
AC Q89553;				
DT 08-DEC-1995 (first entry)				
DE Rat cholesterol 7 alpha-hydroxylase, CYP7, DNA including exons 1-4.				
KW CYP7, cholesterol 7 alpha hydroxylase; transcription factor;				
OS regulatory element; ss.				
FM Rattus sp.				
FH Key				
FT exon				
FT Location/Qualifiers				
FT 3644..3784				
FT /*tag= a				
FT /number= 1				
FT exon				
FT 5400..5640				
FT /*tag= b				
FT /number= 2				
FT exon				
FT 6348..6934				
FT /*tag= c				
FT /number= 3				
FT exon				
FT 7928..7997				
FT /*tag= d				
FT /number= 4				
PN EP-648840-A2.				
PD 19-APR-1995.				
PF 07-OCT-1994; 115856.				
PR 13-OCT-1993; US-135488.				
PR 13-OCT-1993; US-135510.				
PR 13-OCT-1993; US-135511.				
PR 28-JAN-1994; US-187453.				
PA (UYNE-) UNIV NORTHEASTERN OHIO.				
P1 Chiang JYL;				
P1 WPI; 95-148718/20.				
PT Cholesterol 7-hydroxylase (CYP7) gene regulatory elements				
PT Including bile responsive elements, useful for identifying CYP7				
PT transcription factors				
PS Disclosure; Page 48-53; 84pp; English.				
CC Q89553 shows the partial DNA sequence of rat cholesterol 7 alpha-				
CC hydroxylase (CYP7), including exons 1-4 (the whole gene has 5 exons).				
CC CYP7 gene expression is controlled by DNA regulatory elements that				
CC are located within the gene. The location of these regulatory				
CC elements has been identified and they have been isolated. These DNA				
CC fragments are useful in the identification of CYP7 transcription				
CC factors.				
CC				
SO Sequence 7997 BP; 2307 A; 1657 C; 1619 G; 2414 T;				
Query Match				
Best Local Similarity	36.1%;	Score 22;	DB 15;	Length 7997;
Matches	27; Conservative	0;	Mismatches	5; Indels 0; Gaps 0.
Db 4975 tctgttaaaaaacaaacaaacaaacac 5006				
OY 10 TCTTTTCAAAAACAAATATAAACATAAAGC 41				
RESULT 14				
ID T79737 standard; DNA; 7997 BP.				
AC T79737;				
DT 11-DEC-1997 (first entry)				
DE Rat cholesterol 7 alpha-hydroxylase gene.				
KW Cholesterol 7-alpha-hydroxylase; CYP7; HepG2; bile acid; rat; ss.				
OS Rattus sp.				

```

FH Key Location/Qualifiers
FT exon 3644..3784
FT /tag= a
FT /number= 1
FT intron 3785..5399
FT /tag= b
FT /number= 1
FT exon 5400..5640
FT /tag= c
FT /number= 2
FT intron 5641..6347
FT /tag= d
FT /number= 2
FT exon 6348..6934
FT /tag= e
FT /number= 3
FT intron 6935..7927
FT /tag= f
FT /number= 2
FT exon 7928..7997
FT /tag= g
FT /number= 3
FT misc_bind 3300..3422
FT /tag= h
FT /label= BARE
FT /note= "bile acid responsive element"
FT misc_bind 3470..3515
FT /tag= g
FT /label= BARE
FT /note= "bile acid responsive element"
FT misc_bind 3572..3612
FT /tag= h
FT /label= BARE
FT /note= "bile acid responsive element (COUP-TFII binding site)"
FT repeat_region 3581..3591
FT /tag= i
FT /label= DR0
FT /note= "direct nucleotide repeat"
FT enhancer 3300..3379
FT /tag= j
FT enhancer 3404..3515
FT /tag= k
FT misc_signal 3563..3609
FT /tag= l
FT /note= "repressor"
FT misc_signal 3615..3619
FT /tag= m
FT /note= "TATA box"
FT protein_bind 3260..3270
FT /tag= n
FT /note= "AP2 HRE"
FT protein_bind 3467..3474
FT /tag= o
FT protein_bind /note= "HNF3"
FT 3475..3491
FT /tag= p
FT /note= "GRE"
FT protein_bind 3501..3506
FT /tag= q
FT /note= "HNF4"
FT protein_bind 3507..3513
FT /tag= r
FT /note= "HRE"
FT protein_bind 3521..3525
FT /tag= s
FT /note= "HRE"
FT protein_bind 3556..3562
FT /tag= t
FT /note= "HNF3"
FT protein_bind 3563..3574
FT /tag= u
FT /note= "TGT/HNF3"

```

```

FT protein_bind 3573..3577
FT /tag= v
FT /note= "HRE"
FT protein_bind 3584..3589
FT /tag= w
FT /note= "HRE"
FT tata_signal 3615..3619
FT /tag= x
PN WO9720044-A2.
PD 05-JUN-1997.
PE 26-NOV-1996; U18552.
PR 27-NOV-1995; US-562985.
PA (UYNE-) UNIV NORTHEASTERN OHIO COLLEGE MEDICINE.
PI Chiang JYL, Stroup D;
DR WPI: 97-310600/28.
PT Stable HepG2 cells containing regulator of human cholesterol
PT 7alpha-hydroxylase gene - useful in assays for agents that modulate
PT expression of this gene, potentially used for treating defects in
PT bile acid synthesis or cholesterol metabolism
PS Disclosure, Page 27-31; 54pp; English.
CC This rat genomic DNA sequence includes exons that code for
CC cholesterol 7-alpha-hydroxylase (CYP7). CYP7 minigenes, esp.
CC derived from the regulatory region of the human CYP7 gene (see
CC T79736), can be used to transfect HepG2 cells. The regulatory
CC region preferably includes the bile acid responsive elements
CC (BAE) and the DR0 repeat. The cells are used in assays for
CC detecting agents that inhibit or stimulate CYP7 expression, i.e.
CC potential therapeutic agents for treating defects in bile acid
CC synthesis and cholesterol metabolism. The cells closely mimic
CC the physiological responses of CYP7 expression to bile acids and
CC allow large-scale automated screening.
SQ Sequence 7997 BP; 2307 A; 1657 C; 1619 G; 2414 T;

```

```

Query Match 36.1%; Score 22; DB 34; Length 7997;
Best Local Similarity 84.4%; Pred. No. 7.94e+00;
Matches 27; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
DB 4975 tctgttaaaacaaacaaacaaacaaac 5006
Oy 10 TCTTTCAAAACAAATTAACATAAGAC 41

```

```

RESULT 15
ID T65071 standard; DNA: 344 BP.
AC T65071;
DT 29-JAN-1998 (first entry)
DE Canine genomic microsatellite DNA.
KW canine: microsatellite; parentage testing; forensic test; breeding; dog;
KW PCR; LCR; primer; genotype; genetic diversity; polymorphism; ds.
OS Canis sp.
FH Key Location/Qualifiers
FT misc_feature 1..41
FT /tag= a
FT /note= "unique flanking sequence"
FT satellite 42..120
FT /tag= b
FT /rpt_type= TANDEM
FT misc_feature 121..344
FT /tag= c
FT /note= "unique flanking sequence"
PD WO9713876-A1.
PD 17-APR-1997.
PE 27-SEP-1996; U15556.
PF 28-SEP-1995; US-004469.
PA (PEZO-) PE ZOOGEN.
PI Dvorak J, Halverson J, Stevenson T;
DR WPI: 97-235910/21.
PT Canine genotyping using micro-satellite length polymorphism - useful
PT for distinguishing individuals, testing of parentage and analysis of
PT relatedness
PS Claim 10; Page 26; 51pp; English.
CC This canine genomic DNA sequence has internal microsatellite repeats
CC that are flanked by unique flanking regions. The flanking sequences

```


(PL)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

Mpsrch_run n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Tue Dec 22 04:58:18 1998; MasPar time 143.81 Seconds
Tabular output not generated. 950.110 Million cell updates/sec

```

```

Title:                >US-08-530-112A-1449
Description:          (1-62) from US08530112A.seq
Perfect score:       61
N.A. Sequence:
  Comp:              1  GATCCCATCTCTTTCAAAA.....GCAGGACAGCTGANGGAA 62
                   CTGAGGGGTAGAAAAAGTTT.....CGTTCTCTGCACTNCCTTT

```

```
Scoring table:  TABLE default
                  Gap 6
Nmatch  STD :   Dbase 0; Query 0
```

```
Searched: 567134 seqs, 1101898692 bases x 2
Post-processing: Minimum Match 0%
Listing first 45 summaries
```

```
Database:
emb155
1:em_ba 2:em_fun 3:em_htg 4:em_huml 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_vi

Database:
genbank107
15:gb_ba1 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_om 20:gb_ov
21:gb_pat 22:gb_ph 23:gb_pl1 24:gb_pl2 25:gb_pr1
26:gb_pr2 27:gb_pr3 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
32:gb_un 33:gb_vi
```

Statistics: Mean 8.776; Variance 7.778; scale 1.128

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
C 1	46	75.4	113341	25	HUAC002425	Homo sapiens Chromosom	2.50e-07
C 2	45	75.4	156641	26	HUAC002544	Homo sapiens Chromosom	2.50e-07
C 3	28	45.9	77318	21	I66494	Sequence 14 from paten	2.27e+00
C 4	26	42.6	126150	26	ACD00352	Homo sapiens Chromosom	1.15e+01
C 5	26	42.6	167956	26	AC003046	Homo sapiens xp22 Pacs	1.15e+01
C 6	25	42.6	185371	17	HS625H18	Human DNA sequence ***	1.15e+01
C 7	25	41.0	290	20	GC929487	G. gallus microsatellit	2.55e+01
C 8	25	41.0	295	30	GI8347	SW53517 Eric D. Green	2.55e+01
C 9	25	41.0	405	18	MMMSA71	M.musculus (clone D19p	2.55e+01
C 10	25	41.0	7038	28	PF6P19A	Plasmodium falciparum	2.55e+01
C 11	25	41.0	37198	26	AC00526	Homo sapiens chromosom	2.55e+01
C 12	25	41.0	59012	26	HSABIC62	Human abl gene, intron	2.55e+01
C 13	25	41.0	68303	26	HS92M18	Human DNA sequence fto	2.55e+01

C	14	25	41.0	103349	17	AC005800	*** SEQUENCING IN PROG	2.55e+01
C	15	25	41.0	114411	26	HUAC0002551	Human chromosome 16 BA	2.55e+01
C	16	25	41.0	117939	27	HS211A9	Human DNA sequence fto	2.55e+01
C	17	25	41.0	136292	17	HS8B82	Human DNA sequence ***	2.55e+01
C	18	25	41.0	143747	26	HS370M22	Human DNA sequence fto	2.55e+01
C	19	25	41.0	144349	17	HS20208	Human DNA sequence ***	2.55e+01
C	20	25	41.0	151166	17	AC005520	*** SEQUENCING IN PROG	2.55e+01
C	21	25	41.0	153011	17	AC004826	*** SEQUENCING IN PROG	2.55e+01
C	22	25	41.0	156490	26	AC004052	Homo sapiens Chromosom	2.55e+01
C	23	25	41.0	190728	17	AC005038	*** SEQUENCING IN PROG	2.55e+01
C	24	25	41.0	200000	17	AC005141	*** SEQUENCING IN PROG	2.55e+01
C	25	25	41.0	200000	17	AC004618	*** SEQUENCING IN PROG	2.55e+01
C	26	25	41.0	200000	17	AC004624	*** SEQUENCING IN PROG	2.55e+01
C	27	25	41.0	206784	26	HS9317	Homo sapiens DNA sequ	2.55e+01
C	28	24	39.3	6771	26	AG001460	Homo sapiens genomic D	5.58e+01
C	29	24	39.3	2745	27	HS034919	Human white homolog (w	5.58e+01
C	30	24	39.3	3041	26	HUMRSC192	Human mRNA for KIAA00	5.58e+01
C	31	24	39.3	4736	25	HUM16DC86c	Caenorhabditis elegans	5.58e+01
C	32	24	39.3	30758	17	CEH05014	Homo sapiens (subclone	5.58e+01
C	33	24	39.3	34595	23	SPAC27E2	S.pombe chromosome I c	5.58e+01
C	34	24	39.3	41823	26	HSN69F4	Human DNA sequence fto	5.58e+01
C	35	24	39.3	79635	26	AC004650	Homo sapiens chromosom	5.58e+01
C	36	24	39.3	113625	17	AC005163	*** SEQUENCING IN PROG	5.58e+01
C	37	24	39.3	114495	17	HS112K5	Human DNA sequence ***	5.58e+01
C	38	24	39.3	122735	26	AC003954	Homo sapiens chromosom	5.58e+01
C	39	24	39.3	132493	26	AC004583	Human Chromosome 15q11	5.58e+01
C	40	24	39.3	138445	17	AC002979	*** SEQUENCING IN PROG	5.58e+01
C	41	24	39.3	141132	17	HS192P9	Human DNA sequence ***	5.58e+01
C	42	24	39.3	155622	17	AC004839	*** SEQUENCING IN PROG	5.58e+01
C	43	24	39.3	178615	17	HS742C19	Human DNA sequence ***	5.58e+01
C	44	24	39.3	200349	27	HSU85195	Homo sapiens BAC139, C	5.58e+01
C	45	24	39.3	205215	17	HS150C2	Human DNA sequence ***	5.58e+01

ALIGNMENTS

RESULT	1			PRI	10-MAR-1998
LOCUS			DNA		
DEFINITION	Homo sapiens Chromosome 16 BAC clone CIT987SK-A-575C2, complete sequence.				
ACCESSION	AC002425				
NID	g2947051				
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;				
	Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE	1 (bases 1 to 113241)				
AUTHORS	Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J., Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.				
TITLE	Homo sapiens Chromosome 16 BAC clone CIT987SK-A-575C2				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 113241)				
AUTHORS	Adams,M.D.				
TITLE	Direct Submission				
JOURNAL	Submitted (15-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA				
REFERENCE	3 (bases 1 to 113241)				
AUTHORS	Adams,M.D.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-MAR-1998) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA				
REFERENCE	On Mar 10, 1998 this sequence version replaced g1:2924794.				
AUTHORS	Address all correspondence to: Mark Adams The Institute for Genomic Research 9712 Medical Center Dr,				
JOURNAL	Rockville, MD 20850, USA e-mail address: mdadams@tigr.org. The bac location is on chromosome . The orientation of the sequence is from SP6 end to T7 end. Genes were identified by a combination of five methods including: XGRAIL (available by anonymous ftp from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of Washington), Genscan (Chris Burge,				
COMMENT	http://genome.stanford.edu/~chris/GENSCANv.html) searches of the complete sequence against a peptide database, and the Human gene				

BASE COUNT	1944 a	1491 c	1486 g	1929 t	368 others
ORIGIN	/organism="unknown"				
Query Match	45.9%	Score 28;	DB 21;	Length 7218;	
Best Local Similarity	2.1%	Pred. No. 2,27e+00;			
Matches	1;	Conservative	36;	Mismatches 10;	Indels 0; Gaps 0;
Db	1389	YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY	1435		
Cp	62	TTTCCTCAGTCTCTGACGCTTATGTATTGTTTGTG	16		
RESULT	4				
LOCUS	AC005152	126150 bp	DNA	PRI	27-JUN-1998
DEFINITION	Homo sapiens chromosome 17, clone hRPC.1037_O.7, complete sequence.				
ACCESSION	AC005152				
NID	93264565				
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;				
	Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 126150)				
AUTHORS	Birken,B., Fasmann,K., Linton,L., Nusbaum,C. and Lander,E.				
TITLE	Unpublished				
JOURNAL	2 (bases 1 to 126150)				
REFERENCE	Birken,B., Fasmann,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatn,C., Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E., Devon,K., Dewar,K., Donelan,L., Etemadl,S., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacocot,L., Kann,L., Macdonald,P., Marquis,N., McKwan,P., McGurk,A., McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nachman,A., Nahf,R., Naylor,D., O'Connor,I., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R., Stange-Thomann,N., Stiliwell,J., Stojanovic,N., Stone,C., Strickland,C., Subramanian,A., Torrella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-JUN-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
REFERENCE	3 (bases 1 to 126150)				
AUTHORS	Birken,B., Fasmann,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatn,C., Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E., Devon,K., Dewar,K., Donelan,L., Etemadl,S., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Gensheimer,S., Geraiery,K., Gilmartin,T., Grant,G., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacocot,L., Kann,L., Macdonald,P., Marquis,N., McKwan,P., McGurk,A., McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nachman,A., Nahf,R., Naylor,D., O'Connor,I., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R., Stange-Thomann,N., Stiliwell,J., Stojanovic,N., Stone,C., Strickland,C., Subramanian,A., Torrella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-JUN-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
REFERENCE	3 (bases 1 to 126150)				
AUTHORS	Birken,B., Fasmann,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatn,C., Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E., Devon,K., Dewar,K., Donelan,L., Etemadl,S., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Gensheimer,S., Geraiery,K., Gilmartin,T., Grant,G., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacocot,L., Kann,L., Macdonald,P., Marquis,N., McKwan,P., McGurk,A., McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nachman,A., Nahf,R., Naylor,D., O'Connor,I., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R., Stange-Thomann,N., Stiliwell,J., Stojanovic,N., Stone,C., Strickland,C., Subramanian,A., Torrella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-JUN-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
REFERENCE	3 (bases 1 to 126150)				
AUTHORS	Birken,B., Fasmann,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatn,C., Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E., Devon,K., Dewar,K., Donelan,L., Etemadl,S., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Gensheimer,S., Geraiery,K., Gilmartin,T., Grant,G., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacocot,L., Kann,L., Macdonald,P., Marquis,N., McKwan,P., McGurk,A., McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nachman,A., Nahf,R., Naylor,D., O'Connor,I., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R., Stange-Thomann,N., Stiliwell,J., Stojanovic,N., Stone,C., Strickland,C., Subramanian,A., Torrella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-JUN-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
REFERENCE	3 (bases 1 to 126150)				
AUTHORS	Birken,B., Fasmann,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatn,C., Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E., Devon,K., Dewar,K., Donelan,L., Etemadl,S., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Gensheimer,S., Geraiery,K., Gilmartin,T., Grant,G., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacocot,L., Kann,L., Macdonald,P., Marquis,N., McKwan,P., McGurk,A., McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nachman,A., Nahf,R., Naylor,D., O'Connor,I., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R., Stange-Thomann,N., Stiliwell,J., Stojanovic,N., Stone,C., Strickland,C., Subramanian,A., Torrella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-JUN-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
REFERENCE	3 (bases 1 to 126150)				
AUTHORS	Birken,B., Fasmann,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatn,C., Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E., Devon,K., Dewar,K., Donelan,L., Etemadl,S., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Gensheimer,S., Geraiery,K., Gilmartin,T., Grant,G., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacocot,L., Kann,L., Macdonald,P., Marquis,N., McKwan,P., McGurk,A., McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nachman,A., Nahf,R., Naylor,D., O'Connor,I., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R., Stange-Thomann,N., Stiliwell,J., Stojanovic,N., Stone,C., Strickland,C., Subramanian,A., Torrella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-JUN-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
REFERENCE	3 (bases 1 to 126150)				
AUTHORS	Birken,B., Fasmann,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatn,C., Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E., Devon,K., Dewar,K., Donelan,L., Etemadl,S., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Gensheimer,S., Geraiery,K., Gilmartin,T., Grant,G., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacocot,L., Kann,L., Macdonald,P., Marquis,N., McKwan,P., McGurk,A., McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nachman,A., Nahf,R., Naylor,D., O'Connor,I., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R., Stange-Thomann,N., Stiliwell,J., Stojanovic,N., Stone,C.,				

Verinebrata: Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
 Murinae; Mus.
 1 (bases 1 to 405)
 Baron, B.
 Direct Submission
 Submitted (12-JUN-1992) B. Baron, Institut Pasteur, 25 Rue du
 Docteur roix, 75724 Paris Cedex, FRANCE
 (bases 1 to 405)

CD:

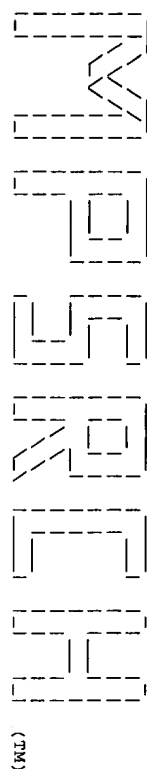
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Alternative splicing of RNAs transcribed from the human abl gene and from the bcr-abl fused gene Cell 47 (2), 277-284 (1986) 6 (sites) Bernards,A., Rubin,C.M., Westbrook,C.A., Paskind,M. and Baltimore,D. The first intron in the human c-abl gene is at least 200 kilobases long and is a target for translocations in chronic myelogenous leukemia Mol. Cell. Biol. 7 (9), 3231-3236 (1987) 8038877 7 (sites) Mount,S.M. A catalogue of splice junction sequences Nucleic Acids Res. 10 (2), 459-472 (1982) 82150208 8 (sites) Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M., Uterback,T.R., Nagle,D.W., Fields,C. and Venter,J.C. Sequence identification of 2,375 human brain genes Nature 355 (6361), 632-634 (1992) 92168112 9 (sites) Zhu,Q.S., Heisterkamp,N. and Groffen,J. Characterization of the human ABL promoter regions Oncogene 5 (6), 885-891 (1990) 90295283 10 (sites) Chen,S.J., Chen,Z., Font,M.P., d'Auriol,L., Larsen,C.J. and Berger,R. Structural alterations of the BCR and ABL genes in P11 positive acute leukemias with rearrangements in the BCR gene first intron: further evidence implicating Alu sequences in the chromosome translocation Nucleic Acids Res. 17 (19), 7631-7642 (1989) 90016849 11 (sites) Tatusov,R. and Lipman,D.J. Using local similarities for pattern detection in nucleotide/protein sequences Unpublished 12 (sites) Groffen,J., Stephenson,J.R., Heisterkamp,N., de Klein,A., Bartem,C.R. and Grosveld,G. Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22 Cell 36 (1), 93-99 (1984) 84106827 13 (bases 1 to 59012) Chisoe,S.L. Sequence of the human abl and bcr genes Thesis (1994) University of Oklahoma 14 (bases 1 to 59012) Roe,B.A. Direct Submission Submitted (08-MAR-1994) Bruce A. Roe, Dept. of Chem. and Biochem., University of Oklahoma, 620 Parrington Oval, Rm 208, Norman, OK 73019, USA	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE Intron	FEATURES SOURCE SOURCE SOURCE
JOURNAL MEDLINE REFERENCE AUTHORS					

```
misc.feature      /citantion=[11]
                  /rpt_family="Alu"
                  18919..18922
/gene="ABL"
/note="breakpoint region: 2 bp from chromosome 9 occurs
in both 9q+ and 22q- sequences at the breakpoint; patient
SA with ANLL (acute non-lymphoblastic leukemia)"
/citantion=[10]
/complement=((19309..19327)..(19636..19663))
/citantion=[11]
/rpt_family="Alu"
repeat_region     20869..21166
                  /citantion=[11]

FEATURES
  source
    1..68903
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="13"
    /map="13q12-13"
    /clone="92M18"
    /clone_lib="RPC1-1"
    94..213
    /note="MIR repeat: matches 201..78 of consensus"
    1633..1743
    /note="AluX repeat: matches 1..110 of consensus;
incomplete repeat"
```

```
repeat_region     1840..1986
/note="L1MD2 repeat: matches 1077..925 of consensus"
repeat_region     1936..2987
/note="MER42c repeat: matches 1303..182 of consensus"
repeat_region     3259..3410
/note="L1MD2 repeat: matches 675..513 of consensus"
repeat_region     3349..3715
/note="L1PA11 repeat: matches 573..226 of consensus"
CDS
join(<3919..4161,6129..6281,7393..8001)
/codon_start=1
/product="BRCA2"
/db_xref="PID:e332399"
/db_xref="PID:g2315186"
/translation="LAPFVLSDECYNILAIKFWIDLNEDIIKPHMLIAASNLQWRE
SKSGILTLFAGDFSVFSASPKRGHGFQETFMKNTVENIDILCNAENKLMHILAND
PKMSPTKDCSTGYPTAOIIPGTGNKILVSMSSPCEIYIOSPLSCAKRKYSTPV
SAOMTSKSGKEKEIDDOCKNRKRALDFLSRLPPLPVSPICFVSPAACKQPPR
SCGTVEPIPKKEKELNSPQMPFRKFNISLSESNISDELAINTALALSGSTGEK
QFISVSESTRAP155EDYLRKRCCTSLIKESSQASVEECKNRKQDTITTKYI
4651..4876
/note="MIR repeat: matches 262..2 of consensus"
5168..5199
/note="16 copies of 2 mer 88 & conserved"
5201..3501
/note="AluX repeat: matches 302..4 of consensus"
6551..6583
/note="3 copies of 11 mer 97 & conserved"
6683..6723
/note="MIR repeat: matches 200..160 of consensus"
6729..6805
/note="MER6 repeat: matches 863..785 of consensus"
6868..6933
/note="MER6 repeat: matches 797..862 of consensus"
6939..7047
/note="MIR repeat: matches 151..40 of consensus"
7146..7510
/note="match: 5' EST H47777 clone 193736"
8250..8383
/note="Alu20 repeat: matches 1..133 of consensus;
incomplete repeat"
9018..9316
/note="AluX repeat: matches 298..1 of consensus"
<10894..11469
/note="match: multiple ESTs; match: F03825 NB0918 Z38606
F02671 R41970 R61510"
14459..14760
/note="AluX repeat: matches 1..302 of consensus"
15125..15432
/note="AluY repeat: matches 301..1 of consensus"
15689..15993
/note="AluX repeat: matches 303..1 of consensus"
18730..18758
/note="MIR repeat: matches 252..224 of consensus"
21408..21565
/note="MER42c repeat: matches 1537..1365 of consensus"
21566..21880
/note="AluX repeat: matches 1..301 of consensus"
21883..21994
/note="MER42c repeat: matches 1373..1262 of consensus"
22155..22446
/note="AluX repeat: matches 290..1 of consensus"
22592..22910
/note="MER42c repeat: matches 1036..706 of consensus"
23099..23400
/note="AluX repeat: matches 302..1 of consensus"
25171..25297
/note="MIR2 repeat: matches 15..146 of consensus"
25400..25543
/note="MIR repeat: matches 258..117 of consensus"
25848..25979
/note="MIR2 repeat: matches 15..146 of consensus"
26055..26332
```


repeat_region	complement(1753. .1898)	repeat_region	/rpt_family="MIR"
repeat_region	/rpt_family="Alu"	repeat_region	complement(27246. .27508)
repeat_region	complement(1964. .2360)	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	complement(27551. .27814)
repeat_region	2499. .2616	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="MER39"	repeat_region	28527. .28821
repeat_region	complement(2825. .3216)	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	complement(29645. .30460)
repeat_region	complement(3585. .3876)	repeat_region	/rpt_family="LTR12"
repeat_region	/rpt_family="Alu"	repeat_region	complement(32175. .32461)
repeat_region	complement(3967. .4263)	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	complement(32632. .32776)
repeat_region	complement(4285. .4494)	repeat_region	/rpt_family="MIR"
repeat_region	/rpt_family="Alu"	repeat_region	32841. .33130
repeat_region	4520. .4822	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	complement(33787. .34077)
repeat_region	5303. .5471	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	34198. .34325
repeat_region	5630. .5843	repeat_region	/rpt_family="MERS"
repeat_region	/rpt_family="Alu"	repeat_region	34634. .34910
repeat_region	complement(6308. .6905)	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	34982. .35257
repeat_region	complement(7105. .7403)	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	complement(40302. .40591)
repeat_region	8844. .9111	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	42724. .42966
repeat_region	complement(9747. .10026)	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	44451. .44737
repeat_region	10173. .10450	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	46054. .46326
repeat_region	complement(10484. .10753)	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	48721. .49012
repeat_region	complement(11003. .11267)	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	complement(49872. .50160)
repeat_region	11443. .12359	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	50370. .50475
repeat_region	complement(12431. .12713)	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	50578. .50679
repeat_region	complement(13128. .13423)	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	complement(50687. .50883)
repeat_region	13897. .14677	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	complement(50987. .51092)
repeat_region	15445. .15731	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="Alu"	repeat_region	51275. .51922
repeat_region	15749. .15859	repeat_region	/rpt_family="Alu"
repeat_region	complement(15926. .16200)	repeat_region	complement(53220. .53389)
repeat_region	/rpt_family="Alu"	repeat_region	/rpt_family="Alu"
repeat_region	complement(16566. .16854)	repeat_region	complement(53610. .53728)
repeat_region	/rpt_family="Alu"	repeat_region	/rpt_family="Alu"
repeat_region	17322. .17717	repeat_region	complement(54332. .54604)
repeat_region	/rpt_family="Alu"	repeat_region	/rpt_family="Alu"
repeat_region	18108. .18381	repeat_region	complement(55229. .55514)
repeat_region	/rpt_family="Alu"	repeat_region	/rpt_family="Alu"
repeat_region	complement(18805. .19078)	repeat_region	complement(55712. .55992)
repeat_region	/rpt_family="Alu"	repeat_region	/rpt_family="Alu"
repeat_region	20080. .20356	repeat_region	complement(57232. .57691)
repeat_region	/rpt_family="Alu"	repeat_region	/rpt_family="Alu"
repeat_region	21406. .21683	repeat_region	complement(58971. .59260)
repeat_region	/rpt_family="Alu"	repeat_region	/rpt_family="Alu"
repeat_region	22565. .22858	repeat_region	complement(60913. .61195)
repeat_region	/rpt_family="Alu"	repeat_region	/rpt_family="Alu"
repeat_region	23472. .23753	repeat_region	62978. .63269
repeat_region	/rpt_family="Alu"	repeat_region	/rpt_family="Alu"
repeat_region	complement(24602. .24760)	repeat_region	complement(63391. .63433)
repeat_region	/rpt_family="Alu"	repeat_region	/rpt_family="Alu"
repeat_region	24837. .25136	repeat_region	64009. .64313
repeat_region	/rpt_family="Alu"	repeat_region	/rpt_family="Alu"
repeat_region	complement(25556. .26020)	repeat_region	64464. .64759
repeat_region	/rpt_family="Alu"	repeat_region	/rpt_family="Alu"
repeat_region	complement(26467. .26675)	repeat_region	complement(67417. .67691)
repeat_region	/rpt_family="Alu"	repeat_region	/rpt_family="Alu"
repeat_region	complement(27092. .27128)	repeat_region	complement(68441. .68725)



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Run on: Tue Dec 22 05:33:30 1998: MasPar time 504.79 Seconds
Tabular output not generated. 1370.839 Million cell updates/sec

Title: >US-08-530-112A-4937
Description: (1-314) from US08530112A.seq
Perfect Score: 298
N.A. Sequence: 1 GATCCTCTCTCCAGGAGT.....CCTNCNTGTGCGCCCTNTTN 314
Comp: CTAGGAGAGAGGGTCCCTTA.....GGANGNACACCGGGAGANAN

Scoring table: TABLE default
Gap 6

Match STD : Dbase 0; Query 0

Searched: 567134 seqs, 1101898692 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

emb15
1:em_ba 2:em_fun 3:em_htg 4:em_hum1 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_vi
genbank107
15:gb_ba1 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_om 20:gb_ov
21:gb_pat 22:gb_ph 23:gb_pl1 24:gb_pl2 25:gb_pl3
26:gb_pr2 27:gb_pr3 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
32:gb_un 33:gb_vi

Statistics: Mean 9.801; Variance 4.588; scale 2.136

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description	Pred. No.
1	278	93.3	5635	26	AF051782	Homo sapiens diaphanous	1.46e-215		
2	278	93.3	68067	26	AC005366	Homo sapiens Chromosom	1.46e-215		
3	32	10.7	216021	26	HUAC004787	Homo sapiens Chromosom	6.17e-06		
4	30	10.1	215	21	128278	Sequence 5 from patent	1.25e-04		
5	28	9.4	74371	26	AC005369	Homo sapiens Chromosom	2.34e-03		
6	26	8.7	215	21	128278	Sequence 5 from patent	3.98e-02		
7	26	8.7	7218	21	156494	Sequence 14 from patent	3.98e-02		
8	26	8.7	10772	18	AF012089	Drosophila melanogaste	3.98e-02		
9	25	8.4	10772	18	AF012089	Drosophila melanogaste	1.57e-01		
10	25	8.4	74371	26	AC005369	Homo sapiens Chromosom	1.57e-01		
11	24	8.1	8860	18	CEU09277	Caenorhabditis elegans	6.04e-01		
12	24	8.1	35586	18	CEU2C416	Caenorhabditis elegans	6.04e-01		
13	23	7.7	254	30	HSP509B11	H. sapiens flow-sorted	2.24e+00		

RESULT	1	AF051782	5635 bp	MRNA	PRI	11-MAR-1998
LOCUS	1	AF051782	5635 bp	MRNA		
DEFINITION		Homo sapiens diaphanous 1 (HDIA1) mRNA, complete cds.				
ACCESSION		AF051782				
NID		92947237				
KEYWORDS		human.				
SOURCE		human.				
ORGANISM		human.				
REFERENCE		1 (bases 1 to 5635)				
AUTHORS		Leon, P.E., Raventos, H., Lynch, E., Morrow, J. and King, M.C.				
TITLE		The gene for an inherited form of deafness maps to chromosome 5q31				
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 89 (11), 5181-5184 (1992)				
MEDLINE		92279220				
REFERENCE		2 (bases 1 to 5635)				
AUTHORS		Lynch, E.D., Lee, M.K., Morrow, J.E., Welch, P.L., Leon, P.E. and King, M.C.				
TITLE		Nonsyndromic deafness DFNA1 associated with mutation of a human				
JOURNAL		Science 278 (5341), 1315-1318 (1997)				
MEDLINE		98028756				
REFERENCE		3 (bases 1 to 5635)				
AUTHORS		Lynch, E.D., Lee, M.K. and King, M.C.				
TITLE		Direct Submision				
JOURNAL		Submitted (06-FEB-1998) Medical Genetics, University of Washington,				
FEATURES		Box 357720, Seattle, WA 98195-7720, USA				
source		Location/Qualifiers				
gene		1..5635				
		/organism="Homo sapiens"				
		/db_xref="taxon:9606"				
		/Chromosome="5"				
		/map="5q31"				
		1..5635				
		/note="mutation associated with nonsyndromic deafness				
		DFNA1"				
		/gene="HDIA1"				

ALIGNMENTS

B. napus mRNA for 3-iso
H. sapiens mRNA for pro
Homo sapiens Chromosom
Zea mays beta-D-glucos
Oxytricha fallax 57kd
Human aromatase cytoch
Mus musculus RNA helic
Methanococcus jannasch
Human (lambda) DNA for
Homo sapiens PAC clone
Human DNA sequence ***
Homo sapiens DNA sequ
Homo sapiens Chromosom
Homo sapiens Chromosom
Human DNA sequence fto
Homo sapiens Chromosom
Human DNA sequence ***
Sequence 145 from pate
Sequence 143 from pate
human STS WI-610.
Cavia porcellus P2X2 r
Rattus norvegicus P2X2
Rattus norvegicus mRNA
Bos taurus myosin X, c
Theraps aquaticus ther
*** SEQUENCING IN PROG
Toxoplasma gondii Chlo
Arabidopsis thaliana D
Mus musculus Chromosom
Homo sapiens Chromosom
Caenorhabditis elegans
Human DNA sequence fto


```

repeat_region complement(20866. .20968)
/rpt_family="MIR"
repeat_region 24489. .24509
/note="(T)21"
/rpt_type=tandem
/rpt_unit="r"
repeat_region complement(24492. .24781)
/rpt_family="Alu"
complement(25029. .25230)
/rpt_family="MIR"
misc_feature complement(25473. .25561)
/note="GRAIL 2 excellent exon, frame 0"
repeat_region complement(25807. .25992)
/rpt_family="Alu"
complement(27340. .27635)
/rpt_family="Alu"
repeat_region 27852. .28087
/rpt_family="Alu"
28491. .28780
/rpt_family="MER42"
repeat_region complement(29213. .29352)
/rpt_family="MIR"
29620. .29657
/note="(GT)19"
/rpt_type=tandem
/rpt_unit="CT"
29791. .30070
/rpt_family="Alu"
repeat_region 30445. .30708
/rpt_family="Alu"
30722. .30761
/note="(AATRA)10"
/rpt_type=tandem
/rpt_unit="AATRA"
30961. .31253
/rpt_family="Alu"
31337. .31258
/note="(A)22"
/rpt_type=tandem
/rpt_unit="A"
31920. .31971
/note="(CTTC)13"
/rpt_type=tandem
/rpt_unit="CTTC"
repeat_region complement(32017. .32286)
/rpt_family="Alu"
33109. .33403
/rpt_family="Alu"
complement(33811. .34024)
/rpt_family="Alu"
/rpt_family="Alu"
complement(39067. .41652)
/note="GRAIL 2 excellent exon, frame 0"
misc_feature complement(43155. .45770)
/note="GRAIL 2 excellent exon, frame 1"
complement(47851. .48154)
/rpt_family="Alu"
complement(48839. .48998)
/rpt_family="Alu"
49009. .49030
/note="(T)22"
/rpt_type=tandem
/rpt_unit="r"
repeat_region complement(49013. .49425)
/rpt_family="Alu"
complement(49876. .50143)
/rpt_family="Alu"
complement(50519. .50839)
/rpt_family="Alu"
complement(52236. .54827)
/note="GRAIL 2 excellent exon, frame 1"
misc_feature complement(55687. .55781)
/note="GRAIL 2 excellent exon, frame 0"
repeat_region 55506. .56645

```

```

repeat_region /rpt_family="MERS"
complement(56856. .57002)
/rpt_family="Alu"
repeat_region complement(57947. .58229)
/rpt_family="Alu"
repeat_region complement(58392. .58683)
/rpt_family="Alu"
59594. .59895
/rpt_family="Alu"
59878. .59897
/note="(A)20"
/rpt_type=tandem
/rpt_unit="A"
59939. .60240
/rpt_family="Alu"
60242. .60355
/rpt_family="MER45"
repeat_region complement(60807. .61083)
/rpt_family="Alu"
complement(61377. .61941)
/rpt_family="Alu"
complement(62074. .62355)
/rpt_family="Alu"
complement(62596. .62924)
/rpt_family="Alu"
62886. .63277
/rpt_family="Alu"
63592. .63888
/rpt_family="Alu"
64602. .64769
/rpt_family="Alu"
complement(64813. .65092)
/rpt_family="Alu"
65891. .66202
/rpt_family="Alu"
66304. .66378
/note="GRAIL 2 excellent exon, frame 1"
complement(66660. .66949)
/rpt_family="Alu"
67199. .67495
/rpt_family="Alu"
67478. .67505
repeat_region

...
Note: remainder of annotations omitted.

Query Match 93.3% Score 278; DB 26; Length 68067;
Best Local Similarity 94.5% Pred. No. 1,46e-215;
Matches 291; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

Db 15181 GATCCTCTCTCCAGGGAATCCGACAGAGAGAACCCCTTCTGTGAGTGGGCCA 15240
|||||
QY 1 GATCCTCTCTCCAGGGAATCCGACAGAGAGAACCCCTTCTGTGAGTGGGCCA 60

Db 15241 GGCTAGAGTACGAGAACTCTTAAGCCACAGAGTTTTATATAATGATATC 15300
|||||
QY 61 GGCTAGAGTACGAGAACTCTTAAGCCACAGAGTTTTATATAATGATATC 120

Db 15301 AACCCAAATGTCAGATGCTTAAGTGCATTCGGGAACTGGGCACAGAGTGGCTTC 15360
|||||
QY 121 AACCCAAATGTCAGATGCTTAAGTGCATTCGGGAACTGGGCACAGAGTGGCTTC 180

Db 15361 ATACACTGTACCCAGCTCTCTTAAGAGAGAGTGGTGGCACT -GAACTGTTTG 15419
|||||
QY 181 ATACACTGTACCCAGCTCTCTTAAGAGAGAGTGGTGGCACTGAACTGTTTG 240

Db 15420 GTGGCCCAACACAGAGAACTGCAATTCGTGGCTTAGGGTATACCTTG -CCCTCTT 15478
|||||
QY 241 GTGGCCCAACACAGAGAACTGCAATTCGTGGCTTAGGGTATACCTTGCCCTMCN 300

Db 15479 TGTGCCCC 15486
|||||
QY 301 TGTGCCCC 308

```

```

RESULT      3
LOCUS       HUAC004787.216021 bp      DNA      PRI      24-JUL-1998
DEFINITION  Homo sapiens Chromosome 16 BAC clone C119875K-A-952F10, complete
              sequence.
ACCESSION   AC004787
            93337381
KEYWORDS    HTG.
SOURCE      human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominoidea; Homo.
REFERENCE   1 (bases 1 to 216021)
AUTHORS    Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
            Homo sapiens Chromosome 16 BAC clone C119875K-A-952F10
            Unpublished
            2 (bases 1 to 216021)
            Adams,M.D. and Loftus,B.J.
            Direct Submission
            Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA, Email:
            bjoftus@tigr.org
            3 (bases 1 to 216021)
            Adams,M.D. and Loftus,B.J.
            Direct Submission
            Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
            On Jul 24, 1998 this sequence version replaced gi:3241936.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr., Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org. The orientation of the sequence is from
            SP6 end to 17 end. Genes were identified by a combination of five
            methods including: XCRail (available by anonymous ftp from
            arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
            Washington), GenScan (Chris Burge,
            http://genomic.stanford.edu/~chris/GENSCANW.html) searches of the
            complete sequence against a peptide database, and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi.html).
            Genes without peptide homology having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAScan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAScan-SE/).
            Location/Qualifiers
            1..216021
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="16"
            /map="16q21-22"
            /clone="A-952F10"
            27765..27872
            /note="7766, STS1-CSRL-2793-uA/CSRL-2793-uZ, Chr. -, Homo
            sapiens"
            /db_xref="dbSTS:G02280"
            73826..73943
            /note="7608, STS1-CSRL-2491-uA/CSRL-2491-uZ, Chr. -, Homo
            sapiens"
            /db_xref="dbSTS:G02122"
            175801..175945
            /note="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"
            /db_xref="dbSTS:G09703"
            175810..175945
            /note="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"
            /db_xref="dbSTS:G09935"
            199463..199572
            /note="9824, WI-3555, Chr. 16, Homo sapiens"
            /db_xref="dbSTS:G04338"
            60960 a 51778 c 49172 g 53987 t 124 others
            BASE COUNT
            ORIGIN
            Query Match      10.7%; Score 32; DB 26; Length 216021;
            Best Local Similarity 8.6%; Pred. No. 6.17e-06;
            Matches 13; Conservative 82; Mismatches 54; Indels 3; Gaps 3;

```

```

Db      1431 BAKKRWCSYTYCCYCSBARRRGKKKMKMSMRARRRGSAKKKKYYY-YVC-YY 1488
Cp      272 AANAAATTGCACCTTCCTGTTGNCACCAACAGTTCAAGTGCCACCCTT 213
Db      1489 YYYYYCMGRAMAAAMYYKKRRSCAMYY-YMKGRARITYYMARRGACRKYMMAYW 1547
Cp      212 CTCCTTTAGAAAGACGCGGGGTACAGTGTATGAAGCAGCAGCCAGTCCAGTCCCA 153
Db      1548 AAWTTWCAAAAMAAATYYMMMYKKMYT 1579
Cp      152 GAATGTCCAGTTAGCATCTGCANATTGCTT 121

RESULT      4
LOCUS       I28278      215 bp      DNA      PAT      30-OCT-1996
DEFINITION  Sequence 5 from patent US 5569830.
ACCESSION   I28278
            91819054
KEYWORDS    .
SOURCE      Unknown.
            ORGANISM
            Unknown.
            Unclassified.
            1 (bases 1 to 215)
            Bennett,A., Labavitch,J.M., Powell,A. and Stoltz,H.
            Plant inhibitors of fungal polygalacturonases and their use to
            control fungal disease
            Patent: US 5569830-A 5 29-OCT-1996;
            Location/Qualifiers
            1..215
            /organism="unknown"
            BASE COUNT      15 a      8 c      25 g      26 t      141 others
            ORIGIN
            Query Match      10.1%; Score 30; DB 21; Length 215;
            Best Local Similarity 20.1%; Pred. No. 1.25e-04;
            Matches 32; Conservative 55; Mismatches 70; Indels 2; Gaps 2;
            Db      49 TYRVNNDGHNKSYSSANNVGNVGAKTHTYHTNSGADSKTYDS-YNASGTSSN 107
            Qy      8 TCTCCAGGAATCGACACAGAGAACCCCTTCTGTTGAGCTGGCCAGCCCTTA 67
            Db      108 G-GTGNNSGADSYSSSTAMTSRNRGTANNVAVDSNMGDASVSDKNTKHAKNASAD 166
            Qy      68 GAGTAGCAGAACTTAAGACACAGAGATTTTTATTAATATATATATCAAGCCAA 127
            Db      167 GKVSGKNNGDRNRRYGTGTSNVSNCGCNKRDVSYA 205
            Qy      128 ATNTGCAATGCTACTGACATCTGTGGGAACTGGCA 166

RESULT      5
LOCUS       AC005369      74371 bp      DNA      PRI      01-AUG-1998
DEFINITION  Homo sapiens chromosome 5, BAC clone 119j3 (DBNL H175), complete
              sequence.
ACCESSION   AC005369
            93367505
KEYWORDS    HTG.
SOURCE      human.
            ORGANISM
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominoidea; Homo.
            1 (bases 1 to 74371)
            Kimmery,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pfluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
            Sequencing of human chromosome 5
            Unpublished
            2 (bases 1 to 74371)
            Rieke,D.O.
            Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
            Unpublished
            3 (bases 1 to 74371)
            JOURNAL
            REFERENCE

```

AUTHORS Kimmery, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
Davis, C.A., Kadner, K., Miguel, T., Pitluck, S., Pollard, M.,
Rojeski, H., Subramanian, S. and Martin, C.H.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
COMMENT Sequence submitted by:
DOE Joint Genome Institute.
FEATURES
source 1. 74371
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="5q"
/clone="11913"
/chromosome="5"
/note="LBNL H175"
893. .1030
repeat_region /rpt_family="Alu"
2295.2438
/rpt_family="Alu"
2818. .2859
/note="(GT) 21"
/rpt_type=tandem
/rpt_unit=GT
join(3246. .3410,3721. .3828)
misc_feature
/standard_name="RUF"
/note="65% & 69% protein identity GenPept:U22377"
3431. .3724
repeat_region /rpt_family="Alu"
3707. .3728
/note="(A) 22"
/rpt_type=tandem
/rpt_unit=A
4366. .4661
repeat_region /rpt_family="Alu"
5327. .5602
/rpt_family="Alu"
6586. .6956
/rpt_family="L1"
6647. .6684
/note="(CA) 19"
/rpt_type=tandem
/rpt_unit=CA
7113. .7373
repeat_region /rpt_family="Alu"
complement(7830. .8185)
repeat_region /standard_name="possible repeat"
8258. .8503
/rpt_family="Alu"
9070. .9387
repeat_region /rpt_family="Alu"
complement(9740. .9845)
repeat_region /rpt_family="MER42"
complement(10440. .11015)
repeat_region /rpt_family="Alu"
11950. .12250
/rpt_family="Alu"
12057. .12085
/note="(A) 29"
/rpt_type=tandem
/rpt_unit=A
12365. .12645
/rpt_family="Alu"
13727. .13750
/note="(AC) 12"
/rpt_type=tandem
/rpt_unit=AC
13783. .14024
/rpt_family="L1"
14175. .14470
/rpt_family="Alu"
repeat_region complement(14906. .15259)

repeat_region /standard_name="possible repeat"
15300. .15613
/rpt_family="Alu"
16671. .16690
/note="(T) 20"
/rpt_type=tandem
/rpt_unit=T
complement(16675. .16977)
repeat_region /rpt_family="Alu"
complement(16993. .17085)
/rpt_family="MER42"
complement(17678. .18276)
repeat_region /rpt_family="Alu"
19305. .19583
/rpt_family="Alu"
19914. .19945
/note="(GTT) 8"
/rpt_type=tandem
/rpt_unit=GTT
complement(19943. .20222)
repeat_region /rpt_family="Alu"
21202. .21496
/rpt_family="Alu"
21736. .22035
/rpt_family="Alu"
22017. .22038
/note="(A) 22"
/rpt_type=tandem
/rpt_unit=A
22321. .22457
/rpt_family="Alu"
complement(22638. .22981)
repeat_region /standard_name="possible repeat"
23022. .23326
/rpt_family="Alu"
23473. .23761
/rpt_family="Alu"
23744. .23767
/note="(A) 24"
/rpt_type=tandem
/rpt_unit=A
complement(24639. .24694)
repeat_region /rpt_family="MER42"
complement(25349. .25713)
repeat_region /rpt_family="Alu"
complement(25727. .26471)
repeat_region /rpt_family="Alu"
27191. .27477
/rpt_family="Alu"
27774. .28057
/rpt_family="Alu"
28040. .28066
/note="(A) 27"
/rpt_type=tandem
/rpt_unit=A
28712. .28930
/rpt_family="MER20"
complement(28769. .28838)
repeat_region /note="GRAIL 2 excellent exon, frame 0"
28987. .29214
/rpt_family="Alu"
join(29485. .29559,29598. .29779)
29495. .29976
/note="95% identity dbSTS:G14532 (SHGC-11312)"
/standard_name="A1027947"
/note="100% identity EST ov84a10.x1"
complement(30401. .30536)
/note="GRAIL 2 excellent exon, frame 2"
complement(30682. .30733)
/rpt_family="MIR"
complement(31573. .31724)
/note="GRAIL 2 excellent exon, frame 1"
complement(32159. .32232)
misc_feature

BASE COUNT	2929 a	2357 c	2282 g	3046 t	158 others
IX"					
Query Match		8.7%	Score 26;	DB 18;	Length 10772;
Best Local Similarity		15.2%;	Pred. No. 3.98e-02;		
Matches	10;	Conservative	36;	Mismatches 20;	Indels 0;
				Gaps 0;	
Db	1694	WVYAWATTMMKKMMTTKMMAMKMTFTYMMKKMYMTSRTTTSAMMMYWTSTWTKYWM	1753		
Cp	129	ATTGGCTGATACATTATATACATTATATAAAACCTGCTGTCTTACAGTTCCTGCTAC	70		
Db	1754	AYAMMK	1759		
Cp	69	TCTTAG	64		
RESULT	9				
LOCUS	AF012089	10772 bp	DNA	INV	05-AUG-1997
DEFINITION	Drosophila melanogaster cysteine proteinase-1 (Cp1) gene, complete cds, and phenylalanyl tRNA synthetase gene, partial cds.				
ACCESSION	AF012089				
NID	g2305220				
KEYWORDS	fruit fly.				
SOURCE	Drosophila melanogaster				
ORGANISM	Eukaryote; mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Preygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 4546 to 4553)				
AUTHORS	Gray,Y.H., Tanaka,M.M. and Syed,J.A.				
TITLE	P-element-induced recombination in Drosophila melanogaster: hybrid element insertion				
JOURNAL	Genetics 144 (4), 1601-1610 (19956)				
MEDLINE	97132596				
REFERENCE	2 (bases 1 to 10772)				
AUTHORS	Gray,Y.H.M., Syed,J.A., Preston,C.R. and Engels,W.R.				
TITLE	Structure of the cysteine proteinase (Cp1) gene of Drosophila melanogaster and associated mutational effects				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 10772)				
AUTHORS	Gray,Y.H.M., Syed,J.A., Preston,C.R. and Engels,W.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (30-JUN-1997) School of Biological Sciences, University of Sydney, Biology A12, Sydney University, NSW 2006, Australia				
FEATURES	Location/Qualifiers				
Source	1..10772				
	/organism="Drosophila melanogaster"				
	/db_xref="taxon:7227"				
mRNA	join(872..1000,2310..2426,6476..6690,6751..7707)				
	/gene="Cp1"				
gene	/product="cysteine protease"				
	872..7707				
exon	/gene="Cp1"				
	872..1000				
intron	/gene="Cp1"				
	/number=1				
exon	1001..2309				
	/gene="Cp1"				
	2310..2426				
	/gene="Cp1"				
	/number=2				
CDS	join(2328..2426,6476..6690,6751..7462)				
	/gene="Cp1"				
	/codon_start=1				
	/product="cysteine proteinase-1"				
	/db_xref="pid:g2305221"				
	/translation="MRTAVLLPLALLAAVAQSPADVYMEWHTKLEHRKNYODET EEFRKIFENKHKIAKQRPAREGCVSKIAVNRKADLLHHEFRDLANGFNLTLRK QLAADESGVTFISPAHVTLRKSYDMRTKGVAVAKDGHGCSWAFFSTGLGEOG HEKSAVLVLSIDONVDCSTKYGNGNGGLMDNMFRIKIDNGIDTEKSPYEALID DCHHPKMGVATDGRFTDIPQODEKKMAEVAATVGPVSAIDAHSHEFQYSGVYN EPCQNDONLDHGVLYVGFDESGEDPWTAKNSMGTWGRKGFIKMLRNENOGCIAS				

repeat_region	complement(16675, .16977)
repeat_region	/rpt_family="Alu"
repeat_region	complement(16993, .17085)
repeat_region	/rpt_family="MER42"
repeat_region	complement(17678, .18276)
repeat_region	/rpt_family="Alu"
repeat_region	19305, .19583
repeat_region	/rpt_family="Alu"
repeat_region	19914, .19945
repeat_region	/note="(GTT)8"
repeat_region	/rpt_type=tandem
repeat_region	/rpt_unit=GTG
repeat_region	complement(19943, .20222)
repeat_region	/rpt_family="Alu"
repeat_region	21202, .21496
repeat_region	/rpt_family="Alu"
repeat_region	21736, .22035
repeat_region	/rpt_family="Alu"
repeat_region	22017, .22038
repeat_region	/note="(A)22"
repeat_region	/rpt_type=tandem
repeat_region	/rpt_unit=A
repeat_region	22321, .22457
repeat_region	/rpt_family="Alu"
repeat_region	complement(22638, .22981)
repeat_region	/standard_name="possible repeat"
repeat_region	23022, .23326
repeat_region	/rpt_family="Alu"
repeat_region	23473, .23761
repeat_region	/rpt_family="Alu"
repeat_region	23744, .23767
repeat_region	/note="(A)24"
repeat_region	/rpt_type=tandem
repeat_region	/rpt_unit=A
repeat_region	complement(24639, .24694)
repeat_region	/rpt_family="MER42"
repeat_region	complement(25349, .25713)
repeat_region	/rpt_family="Alu"
repeat_region	complement(25727, .26471)
repeat_region	/rpt_family="Alu"
repeat_region	27191, .27477
repeat_region	/rpt_family="Alu"
repeat_region	27774, .28057
repeat_region	/rpt_family="Alu"
repeat_region	28040, .28066
repeat_region	/note="(A)27"
repeat_region	/rpt_type=tandem
repeat_region	/rpt_unit=A
repeat_region	28712, .28930
repeat_region	/rpt_family="MER20"
repeat_region	complement(28769, .28838)
repeat_region	/note="GAA1, 2 excellent exon, frame 0"
repeat_region	28987, .29214
repeat_region	/rpt_family="Alu"
repeat_region	join(29485, .29559, 29598, .29779)
repeat_region	/note="95% identity dbSTS:G14522 (SHGC-11312)"
repeat_region	29495, .29976
repeat_region	/standard_name="A1027942"
repeat_region	/note="100% identity EST ov84a10.x1"
repeat_region	complement(30401, .30366)
repeat_region	/note="GAA1, 2 excellent exon, frame 2"
repeat_region	complement(30682, .30733)
repeat_region	/rpt_family="Mir"
repeat_region	complement(31573, .31724)
repeat_region	/note="GAA1, 2 excellent exon, frame 1"
repeat_region	complement(33159, .33232)
repeat_region	/note="GAA1, 2 excellent exon, frame 2"
repeat_region	33388, .32488
repeat_region	/rpt_family="MLT1"
repeat_region	32617, .32908
repeat_region	/rpt_family="Alu"
repeat_region	32877, .33086
repeat_region	/rpt_family="MLT1"

FEATURES	source
repeat_region	complement(33570..33785) /rpt_family="Alu"
misc_feature	complement(34021..34144) /note="GRAIL 2 excellent exon, frame 2"
misc_feature	complement(35238..35331) /note="GRAIL 2 excellent exon, frame 1"
repeat_region	36382..36663 /rpt_family="Alu"
misc_feature	36901..37222 /note="100% identity EST ou55c09.x1"
STS	36901..37164 /standard_name="STSC-9983"
misc_feature	/db_xref="dbSTS:G26554"
misc_feature	complement(37595..37654) /note="GRAIL 2 excellent exon, frame 0"
misc_feature	complement(join(18065..38218,38462..38578,38741..38993,39071..39205,39532..39630,39935..40046,40300..40410,40503..40661,41868..41972,42103..42225,42492..42569,44379..44507))
misc_feature	/standard_name="histidyl-tRNA synthetase"
misc_feature	/note="78%-100% protein identity GenPept:U18937"
misc_feature	complement(38069..38215) /note="GRAIL 2 excellent exon, frame 0"
misc_feature	complement(38462..38578)
..	Note: remainder of annotations omitted.
Query Match	8.4%; Score 25; DB 26; Length 74371;
Best Local Similarity	10.6%; Pred. No. 1.57e-01;
Matches	10; Conservative 48; Mismatches 36; Indels 0; Gaps 0
Db 15909	SRSRQSTCKCKCTSTSCSCSRKSRKGRYRPMWMTCKSSKCMWSYMR 15968
Oy	146 GACATTCGGGGAAGTGGGACACAGAGNCTGCTTATACACTGACCCAGCTCTNTCT 205
Db 15969	MKCYSCSYCCSSGKRKYRCRSMWYTYRSYK 16002
Oy	206 AAAAGAGAAGTGGTGGGCACACTGAACTGTT 239
RESULT 11	
LOCUS	CEU092377 8860 bp DNA INV 23-NOV-1994
DEFINITION	Caenorhabditis elegans Bristol N2 synaptic vesicle acetylcholine transporter (unc-17) gene, complete cds.
ACCESSION	U092377
NID	9501061
KEYWORDS	
SOURCE	Caenorhabditis elegans.
ORGANISM	Caenorhabditis elegans. Eukaryotes; mitochondrial eukaryotes; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditidae; Rhabditidae; Rhabditidae; Rhabditidae.
REFERENCE	1 (sites) Alfonso, A., Grundahl, K., McManus, J. R., Asbury, J. M. and Rand, J. B. Alternative splicing leads to two cholinergic proteins in Caenorhabditis elegans
AUTHORS	J. Mol. Biol. 241 (4), 627-630 (1994).
TITLE	
JOURNAL	9435008
MEDLINE	9435008
REFERENCE	2 (sites) Alfonso, A., Grundahl, K., Duerr, J. S., Han, H. P. and Rand, J. B. The Caenorhabditis elegans unc-17 gene: a putative vesicular acetylcholine transporter
AUTHORS	Science 261 (5121), 617-619 (1993)
TITLE	
JOURNAL	9354294
MEDLINE	9354294
REFERENCE	3 (bases 1 to 8860) Rand, J. B.
AUTHORS	Direct Submission
TITLE	Submitted (27-APR-1994) James B. Rand, Oklahoma Medical Research Foundation, Program in Molecular and Cell Biology, 825 N.E. 13th Street, Oklahoma City, OK 73104, USA
JOURNAL	
MEDLINE	
REFERENCE	1..8860
QUALIFIERS	

```

/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/clone="RM no.88P; RM no.104P; RM no.148P"
/chromosome="IV"
2401..2466
/gene="unc-17"
/note="This is the first exon for both the unc-17
transcript and the cna-1 transcript; see GenBank Accession
Number L08969"
number=1
join(2401..2466,3613..4115,5781..6060,7770..8802)
/gene="unc-17"
/citation=[1]
/citation=[2]
/evidence=experimental
join(2401..2466,3613..4115,5781..6060,7770..8802)
/gene="unc-17"
2467..3612
/number=1
3613..4115
/gene="unc-17"
/number=2
join(3627..4115,5781..6060,7770..8599)
/gene="unc-17"
/citation=[1]
/citation=[2]
/codon_start=1
/function="synaptic vesicle acetylcholine transporter"
/evidence=experimental
/product="UNC-17"
/db_xref="PID:g501062"
/translation="MGFNPVPIINDSEILKADAKKLEQDQNKCVLVIVSIALILD
NMLTVVPIIPKLRDINHVFEEYHETSQLANGTLLVREVGGRINFLDEEEL
GWLFASKALQIFNPSGYIDRVGEISALGCTMFSTAFALGKSGVLEFAR
SLQFGSAGADISGLAMADRTENERSALGALAFISGCIVAPFSGVSLAG
KVPFELISFVCLADAIAMVIMPHRGDSHGKVGOGPMRELNDPILACCGSL
IMANVSIALEPIPTTMMSEMDPTGMIWGVIMLPFPHVIGVYTVYMLARFPH
TWAIVAGLAMEGIACRAIPPTTSVMQVLVPLSVCGIALIDISLIPMLGHLVDTR
VSVGAVYIADISYLAIVAGPIIAGWIVTMGFTALNITFATNTVAPVLELRK
VHSYDTLGAKGDTMEMQLNSSAPAGVNGKPEATTAESYOGMEDQSYONQAIPINH
AVSFQDRPOAEFPAGYDPLNPQW"
4116..5780
/number=2
5781..6060
/gene="unc-17"
/number=3
6061..7769
/number=3
7252..7730
/note="Stem region is 195 nucleotides, with 193/195
complementary nucleotides, loop is 89 nucleotides"
7770..>8801
/gene="unc-17"
/number=4
8802
/gene="unc-17"
/note="See GenBank Accession Number L19621, unc-17 mRNA
sequence"
/citation=[1]
/citation=[2]
/evidence=experimental
8807
/note="See GenBank Accession Number L19621, unc-17 mRNA
sequence"
/citation=[1]
/citation=[2]
/evidence=experimental
8844
/note="See GenBank Accession Number L19621, unc-17 mRNA
sequence"
/citation=[1]
/citation=[2]

```

```

misc-feature
8855..8860
/note="Nucleotides 8855-8860 of this sequence correspond
to nucleotides 1-6 of GenBank Accession Number L08970"
/citation=[1]
/evidence=experimental
/label=SalI_site
BASE COUNT      2751 a      1675 c      1501 g      2931 t      2 others
ORIGIN
Query Match      8.1%; Score 24; DB 18; Length 8860;
Best Local Similarity 74.0%; Pred. No. 6,04e-01;
Matches 37; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Db 2865 AAAATGGCGGAAATTAAGAACACAGAAATTTTCATAAAATTTAAAT 2914
66 AAGAGTAGCAGAACCTAGACACAGAGTTTATTAATATTAATTAAT 115
Qy
RESULT 12
LOCUS      CELZC416      35586 bp      DNA      INV      07-DEC-1997
DEFINITION Caenorhabditis elegans cosmid ZC416.
ACCESSION AF036701
NID 92662609
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans strain-Bristol N2.
Eukaryotae; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditiidae;
Rhabditiina; Rhabditiodea; Rhabditiidae; Pelodierinae; Caenorhabditis.
1 (bases 1 to 35586)
Wilson,R., Almscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fullton,L., Gardner,A., Green,P., Hawkins,T., Haller,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., Mckurray,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Ritken,L., Roopra,A.,
Saunders,D., Showkeen,R., Smaildon,N., Smith,A., Sonhammer,E.,
Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
Vaughan,K., Waterston,R., Watson,A., Welstock,L.,
Wilkinson-Sproat,J. and Wohlman,P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 35586)
AUTHORS Favello,A. and Goela,D.
TITLE The sequence of C. elegans cosmid ZC416
JOURNAL Unpublished (1997)
REFERENCE 3 (bases 1 to 35586)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1997) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: twenematode.wustl.edu and jesusanger.ac.uk

```

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is Y37E11B, 3350 bp overlap; 3' cosmid is C68, 200 bp overlap. Actual start of this cosmid is at base position 1 of CELZC416; actual end is at 35586 of CELZC416

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES

```
source
1..35586
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="ZC416"
4349..6700
/gene="ZC416.4"
join(4349..4387,4431..4522,4742..4871,4913..5028,
5396..5480,5534..5699,5768..6028,6075..6169,6557..6700)
/gene="ZC416.4"
/note="similar to the CDC2/CDC28 subfamily of SER/THR
protein kinases"
/codon_start=1
/evidence=not_experimental
/db_xref="PID:g2662611"
/translation="MNPSPKHOVITTDITPAGAGITPTTTPRYVNLKFTLVAGQIV
MADDVTTQOVAIKKMOQFPLTMSARRAYREPLITLKHPIIRILNFTDLS
SFOEVLVMEFTNHLNHEVIRLRLDHELTSLFVQSLCAIKHLNMGFIHRLKSN
IVNNECVLDEPLDGLARRNVMRSVDYGTYYRAPEVILGLNYSKVDIVSGCI
LAMINHTVLFPGKDIRDQWTOISSVLTPTDDHINOLEPIGAIVYRSPLQHLARFN
EIVDSNLEPETENPKDHLTLHARDLLENMLKINPEERYSVESDALNHPYTMDEVNA
PASEIDDEVNDIDMTLIEKRLITRENOGYOTDHDHFE"
6189..6368
/note="C. elegans telomere-like hexamer repeat"
complement(6888..9666)
/gene="ZC416.6"
complement(join(6888..7029,7079..7183,7235..7794,
8119..8787,9061..9142,9193..9341,9610..9666))
/gene="ZC416.6"
/note="similar to the peptidase family M1 (zinc
metalloproteases)"
/codon_start=1
/evidence=not_experimental
/db_xref="PID:g2662612"
/translation="MIIIGQATLRCRLTDTATKVLVDVDSIRSVSINGVDCDFRIAP
NVYTFGSKMSVYLPPQFOKAGTILQVTVAVGSPDATALQMKKQDQDTRPPLYS
OCQAIHARSIVPCMDTPSVKSYEAETVATGTCMLASIGQSKSDQDTPFYKOP
VAIPSLIAIVGCLERDISDRCVAWAEBSVDKAWMEAREDEDILASAEETAGKYI
WGRYDWVCLPSPFPGFMENPCLFPLPTLADRSIVSYAIEHIASGNGNVTSS
WEHFNINEGTFTIERKICGRIVSEDRQRMANGNTNSIIPRYEOTFTTHQFTLI
QDHTNVDPAFSCVPEKGSALFTLEQKLGSELEFALIKDYLKTFHAQAYDSQW
KDLHYFADQKELVDIVNDFLWNGVMPKPEPNQSGWVECEVLTKMLAADST
VGDITSEDIYAMQPLQOIELLSQLMQHDPIEYKIDALIKLINDOSSEILLMWL
RLCISKECECIKALDEFVNSGRLEKRCRIYDLMQWPIANKKARDVYLKRSQWHP
ITAEMLAKDHIHSTHTVCF"
complement(11607..15328)
/gene="cha-1"
complement(join(11607..11963,12207..12405,12523..12746,
13011..13253,13296..13464,13509..13785,13857..14029,
14073..14214,14745..14794,15278..15328))
/gene="cha-1"
/note="ZC416.8b"
/codon_start=1
/product="C. elegans choline O-acetyltransferase CHA-1
(SW:P32756); ZC416.8a and ZC416.8b are different
alternatively spliced forms of the same gene that both
share a common 5' untranslated exon (Alfonso et al.,
Journal of Molecular Biology 241: 627-630, 1994)"

repeat_region
/note="C. elegans vesicular acetylcholine transporter
UNC-17 (SW:P34711); ZC416.8a and ZC416.8b are different
alternatively spliced forms of the same gene that both
share a common 5' untranslated exon (Alfonso et al.,
Journal of Molecular Biology 241: 627-630, 1994)"
/db_xref="PID:g2662614"
/translation="MGNFVYINRDSIEILAKDAKWLBDQDNKKCVLTVSIALILD
NMLYVAVIPILIKRYLNDHNVOYTERGYNESQLANGVYIAVEGRINFLDEEL
GMFLFWKALQIFVNPFGSGYIDRWYEGEIPMLIGCTMFEIATLALGKSYLLFLAR
SLOGGSASADISGLAMIDRPEENERSALGIALAFISFCILVAPPGSVYSLAG
KVPPLILSFVCLADAIVAVYINPHRGTDSHGKVGSTPMWRLFMDFEITCCGAL
IVANSLALEPTITITWSEMPDIPGHTVGYIWIIPRPHYGLVYVYKMLRAPPH
TVALAVGLAGLINCFALEPTTSVQWOLVPLSFQIALDITSLPLMLGLVTRH
VSVGSVAIADISYSLAVAFGPILAGWVTVWGFTALNIIIFANVYAPVLFILRK
VHSYDILGEFKIQRASESURLSPCLQTFWTLINKFGNMLIMGRKILMLKMLALFE
AVSFQDSRQAGFEPAGYDPLNPQW"
31175..33602
/gene="ZC416.2"
join(31175..31228,31853..31941,33368..33602)
/gene="ZC416.2"
/codon_start=1
/evidence=not_experimental
/db_xref="PID:g2662610"
/translation="MSNAPPYTAQOPPYGNANAPQGSNPYPAGPPQGYIPNQ
PVITWLGFEFKIQRASESURLSPCLQTFWTLINKFGNMLIMGRKILMLKMLALFE
QVVDPEKRYTVLKQIWTILSKI"

gene
11164 a 6603 c 6811 g 11008 t
BASE COUNT
ORIGIN
Query Match 8.1%; Score 24; DB 18; Length 35586;
Best Local Similarity 74.0%; Pred. No. 6.04e-01;
Matches 37; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Db 21737 ATTTAAATTTTGAAGAAATCTGTCTTTATTTCCGCCAATTT 21786
||||| | ||||| ||||| ||||| | ||||| | ||||| | ||||| |
Cp 115 ATTATATCATTTTATTAATAAAGCTGTGIGCTTGTAGATTCCTGCTACTCT 66

RESULT 13
LOCUS HSPD09B11 254 bp DNA STS 21-MAY-1998
DEFINITION H.sapiens flow-sorted chromosome 1 HindIII fragment, SCIPD09B11,
sequence tagged site.
ACCESSION AL009409
NID 92664572
KEYWORDS STS; single read.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
```

REFERENCE 1 (bases 1 to 254)
AUTHORS Gregory,S., Kettleborough,R., Langford,C., Ross,M.T. and Hunt,S.E.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1997) E-mail contact: humquery@esanger.ac.uk
COMMENT Vector: pBS1SK+
Marker stsG29027 (Primer A : GATCCTCATACAGTAGGCG; Primer B : TGAAGCAGAAATGATACCG; amplicmer size : 180 bp) was mapped to chromosome 1 using Radiation Hybrid panel Genebridge 4 (GB4).
FEATURES
source
1..254
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/sex="Female"
/dev_stage="adult"
/tissue_type="EBV lymphoblastoid cell line"
/clone_lib="SC1PE"
/clone="SC1PE09B11"
BASE COUNT 90 a 51 c 39 g 74 t
ORIGIN
Query Match 7.7%; Score 23; DB 30; Length 254;
Best Local Similarity 92.6%; Pred. No. 2.24e+00;
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 24 ATAAAGTATAATTTATGAGCCAA 50
OY 102 ATAAATGATATAATGATCAAGCCAA 128
RESULT 14
LOCUS B.napus 1428 bp RNA PLN 09-MAR-1992
DEFINITION B.napus mRNA for 3-isopropylmalate dehydrogenase.
ACCESSION X59970
NID 917826
KEYWORDS 3-isopropylmalate dehydrogenase.
SOURCE rape.
ORGANISM Brassica napus
Eukaryote; mitochondrial eukaryotes; Viridiplantae;
Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
Magnoliopsida; Caprales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 1428)
AUTHORS Ellestrom,M.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-1991) M. Ellestrom, Dept of Cell Research, Box 596, S - 751 24 Uppsala, SWEDEN
2 (bases 1 to 1428)
REFERENCE 1 (bases 1 to 1428)
AUTHORS Ellestrom,M., Josefsson,L.G., Rask,L. and Ronne,H.
TITLE Cloning of a cDNA for rape chloroplast 3-isopropylmalate dehydrogenase by genetic complementation in yeast
JOURNAL Plant Mol. Biol. 18 (3), 557-566 (1992)
92163020
FEATURES
source
Location/Qualifiers
1..1428
/organism="Brassica napus"
/strain="Svalfs Karat 20516-K"
/db_xref="taxon:3708"
17..118
/note="3-isopropylmalate dehydrogenase; chloroplast targeting signal"
transit_peptide 17..118
CDS
17..1237
/EC_number="1.1.1.85"
/codon_start=1
/product="3-isopropylmalate dehydrogenase"
/db_xref="PID:g17827"
/translation="MAALOTNIRPKPPATRALTKOSSPAPFVRCAASPKKRY
NITLLPGDSIGPEVISIAKNVIAQAGSLGLEFSROEMVGAALDLGVLPLEETVS
AKESDAVLGAIGGKMDKNEHLKPEGLQLRAGLKVFNALRPATVLPQVAST
LKREVAEGVDLAVRELGGIYGVPRGKTNENGEVGYNTENVAAHEDIDIAVAF
ETARRRGKLCISVDKANYADASILMRRTVTAIAAEYDVELSHMTVDNNAQMLVBDPK
QDTITVNNIFGDIILSDASMITGSIGMLPSASLSDSGGLFEPHGSAPDIAQDKA

mat_peptide
/EC_number="1.1.1.85"
/product="3-isopropylmalate dehydrogenase"
polya_signal 1384..1389
polya_site 1410
BASE COUNT 389 a 286 c 364 g 389 t
ORIGIN
Query Match 7.7%; Score 23; DB 23; Length 1428;
Best Local Similarity 71.4%; Pred. No. 2.24e+00;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
Db 430 TGGCTTAAGTCTTGCAATCTGAGACCTGCTACAGTTCTCCACAG 478
Cp 90 TGGCTTAAGTCTTGCAATCTGAGACCTGCTCCAGTCCAGTCAACCANAG 42
RESULT 15
LOCUS HSPCM1 2891 bp RNA PRI 14-NOV-1997
DEFINITION H.sapiens mRNA for protein containing MBD 1.
ACCESSION Y10746
NID 92239125
KEYWORDS methyl-Cpg-binding protein; PCM1 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2891)
AUTHORS Cross,S.H., Meehan,R.R., Nan,X. and Bird,A.
TITLE A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HXK proteins
JOURNAL Nature Genet. 16 (3), 256-259 (1997)
97351508
REFERENCE 2 (bases 1 to 2891)
AUTHORS Cross,S.H.
TITLE Direct Submission
JOURNAL Submitted (23-JAN-1997) S.H. Cross, Edinburgh University, Institute of Cell and Molecular Biology, Darwin Building, Kings Buildings, Mayfield Road, Edinburgh EH9 3JR, UK
Location/Qualifiers
1..2891
/organism="Homo sapiens"
/isolate="Soares retina N2B5HR"
/db_xref="taxon:9606"
/chromosome="18"
/cell_type="retina"
/clone_lib="222272"
/clone_lib="222390"
/dev_stage="adult"
/map="18q12.3-q21.31 mu"
140..1810
/gene="PCM1"
140..1810
/gene="PCM1"
/function="binds methyl-CpGs"
evidence-experimental
/codon_start=1
/product="methyl-Cpg binding protein"
/db_xref="PID:e300065"
/db_xref="PID:g2239126"
/translation="MAEDWLDCCPALGFGWRRREYFRSGATCGRSITYOSPTGDRIR
SKVELTRYLPACDLTLFDFKOGILCYPARAHVAAASRRKRSRPARTRKROVGP
OSGVRKEAPDEKADTDPAASFPAAGCCENGISFSDGTOROLTKTLCKRQV
RIAFNRBOBMKSRKSGVCGVCGCOTGDEGCHGCTCLRRPRLGRLRQWKCVGRICRGH
ARRGGCDSKAARRRGAQPIPPPPSSQSEPEPPRALASPPPEFYICYDEDE
LQPTNRKONKGCACACLRNCGCDDFCDDPKFGSSNQRKRCRMQCLOFAK
RLPSTVNSSESDGASPPYRRRRPSSARHHLGPTLKPTLATRTAQPHTOPTQK
EAGGFTVLPPTGTLVFLRBGASXPVQVPPVAASTBALQEQCSGMSVVALPYK
OEKRDQDEMTGPAVLISPLVYVGCPSKAVDPLGVKQDEPPDEPKDEENKDDSS
KLAEFEAGAGTGVITEISIGTGRPDNAVMIPRSLDKKGRARQO"
BASE COUNT 705 a 784 c 798 g 575 t 29 others

ORIGIN

Query Match	7.7%;	Score 23;	DB 26;	Length 2891;
Best Local Similarity	43.9%;	Pred. No. 2.24e+00;		
Matches	29;	Conservative	12;	Mismatches 25;
			Indels	0;
			Gaps	0;

[illegible]

Search completed: Tue Dec 22 05:42:04 1998
Job time : 514 secs.

(7)

Dd	59	gggywcogagcygcgaayycdhvccgymtttllhyrrmbdnryrdynrsdaawycyrr
QY	162	GGGACGCCAGGNGTSCCTTCATACACTGTACC GCCAGCTCTNTTTAAAGAAGAACTNGTG
Dd	119	rsvkydcynaechddhhvybbbyvnynhnbnnccecbnmhvchnvbnnhmwayrhd
QY	222	GGCACACTGTGAAGTCTTTGGTGTGNCAACAACACAGANGCTGCATTINTNGNTTAGCG
Dd	179	aardavh 185
QY	282	TGATACT 288
 RESULT 3 ID Q51746 standard; cDNA; 91 BP. AC Q51746; DT 31-MAY-1994 (first entry) DE Oligonucleotide probe MK14-A KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis; ss. OS Synthetic. PN EP-571911-A. PD 01-DEC-1993. PE 24-MAY-1993; 108325. PR 26-MAY-1992; US-888651. PA (BECT) BECTON DICKINSON CO. FT Shank DD, Spears PA; RT WPI; 93-378844/48.		
New oligo:nucleotide probes specific for Mycobacteria - used for		

```






PS Claim 3, Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
S0 Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 11.7%; Score 35; DB 9; Length 91;
Best Local Similarity 0.0%; Pred.No. 3,82e-08;
Matches 0; Conservative 41; Mismatches 8; Indels 0; Gaps

Db 12 svhsyyrvhvshhsbhvhvshvshvshvshvshvshvshvshvshvshvshvshvsv 60
CP 89 GGTCTTAGCTTCCTGCTACTCTTAGGCTTGCCCAAGTCAACCANAGA 41
RESULT 4
ID N81164 standard; DNA; 204 BP.
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key location/Qualifiers
FT misc-feature 19..69
FT /*tag-a
FT /function-multiple cloning site
FT primer_bind 187..204
FT /*tag-b
FN EP-285123-A.
FD 05-MAY-1988.
FE 30-MAR-1988; 105163.
FR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Kolvula A, Bamford J, Reinikainen T;
PI WPI: 88-279927/40.
DR Introducing random point mutations into nucleic acids -
FT by prepn of single stranded template, annealing a primer, elongation,
FT misincorporation, completion of molecules and screening.
PS Disclosure; P; English.

```


[illegible][illegible]

Education/Qualifiers
289

source

1. .277

/organism="Homo sapiens"

/note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-Oligo dt adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested lambda ZAP Express."

/db_xref="taxon:9606"

/clone_lib="Fetal heart, Lambda ZAP Express"

/lab_host="E. coli XL1-Blue"

<1. .>277

BASE COUNT 72 a 68 c 71 g 66 t

ORIGIN

Query Match 48.0%; Score 143; DB 22; Length 277;

Best Local Similarity 95.3%; Pred. No. 2.82e-236;

Matches 163; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

Db 109 GATCCTCTCCAGGA-TCCGACACAGAGAACCCCTTCCTGTTGAGCTGGCCA 167

|||

QY 1 GATCCTCTCCAGGAATCCGACACAGAGAACCCCTTCCTGTTGAGCTGGCCA 60

|||

Db 168 GGCCTAAGACTAGCAGACTCTAGACACAGAGTTT-ATAATGTATAATGTATC 226

|||

QY 61 GGCCTAAGACTAGCAGACTCTAGACACAGAGTTTATAATGTATAATGTATC 120

|||

Db 227 AACCCAAATGTGAGATGCTACTGGGACATTTGGGGAGTGGACACAG 277

|||||

QY 121 AACCCAAATGTGAGATGCTACTGGGAGTGGGGAGTGGGGACACAG 170

|||||

RESULT 6

LOCUS AA754459 252 bp mRNA EST 20-JAN-1998

DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII cDNA library Oryza sativa

ACCESSION AA754459

NID 92801165

KEYWORDS EST.

SOURCE rice.

ORGANISM Oryza sativa

Eukaryote; Viridiplantae; Charophyta/Embryophyta group;

Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;

Poales; Poaceae; Oryza.

1 (bases 1 to 252)

Nahm,B.H., Kim,J.R., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,

Kim,M.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,

Lee,M.C. and Eun,M.Y.

Large-scale Sequencing Analysis of ESTs from Rice Immature Seed

Unpublished (1998)

TITLE

JOURNAL

COMMENT

Contact: Eun M.Y.

Department of CytoGenetics

National Inst. of Agri. Sci. and Tech, RDA

Suwon, Kyunggi-do, Korea

Tel: 82 331 290 0301

Fax: 82 331 290 0307

Email: myeun@sun20.asti.re.kr

Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji

University, Yongin, Korea. 449-728 bhnahebloserver.myongji.ac.kr

Seq primer: M13 Reverse Primer.

FEATURES

source

1. .252

/organism="Oryza sativa"

/cultivar="Milyang23"

/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2: XhoI; Directional cDNA library inserted into lambda ZAPII

vector at 5' end with EcoRI and 3' end with Xho I site."

/db_xref="taxon:4530"

/clone="97SN1787"

/clone_lib="Rice Immature Seed Lambda ZAPII cDNA library"

/tissue_type="Immature Seed"

/dev_stage="5 days after pollination"

BASE COUNT 5 a 21 c 12 g 35 t 179 others

ORIGIN

Query Match 17.1%; Score 51; DB 12; Length 252;

Best Local Similarity 7.2%; Pred. No. 1.33e-51;

Matches 16; Conservative 117; Mismatches 86; Indels 4; Gaps 4;

Db 30 VWCVAASHGNWYSHNCTBRGTHCDCKNWNSTWGTWVNBWNSGDWYHBY-BNTKYD 88

|||

CP 312 ANAGGGGCCACANGNAGGGCCAAAAGTATACCCCTAACCCANANANATTCAGACNTCTGT 253

|||

Db 89 VGNHTGCSMRBVTMAHHDYTNGBYNN-DYMMHBMVBTGCTCTTMCBHNNT 147

|||

CP 252 GGTTCGNCACCAACCACTCAAGTGTGCCCCACCACTCTCTTTAGANAGAGCTGG 193

|||

Db 148 KCTASGHTSTVDKSNVTWGTBSYDKSHGWSBVBVYHRTVTRATRTGTCVR 207

|||

CP 192 GGTAC-AGTGATGAAGGACACNCCTGTGCCAGTCCCAAGATGTCAGTATCATCT 134

|||

Db 208 KYCVMMTKRVKRYHVBGCHBTDSKCKTMMTKHVMST 250

|||

CP 133 GCANATTGCTGATACATTTAT-ACATTTATAAAAACTCT 92

|||

RESULT 7

LOCUS AA503247 509 bp mRNA EST 19-AUG-1997

DEFINITION ne45h02.s1 NCI-CGAP_C03 Homo sapiens cDNA clone IMAGE:900339, mRNA

sequence.

ACCESSION AA503247

NID 92238214

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae;

Homo.

1 (bases 1 to 509)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,

M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arraying: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/ILNI at:

www.bio.lnl.gov/db/ftp/image/image.html

Insert Length: 891 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 474.

FEATURES

source

1. .509

/organism="Homo sapiens"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was prepared from 12 pooled bulk tumor samples and primed

with a Not I - oligo(dT) primer. Double-stranded cDNA was

ligated to Eco RI adaptors (Pharmacia), digested with Not

I and cloned into the Not I and Eco RI sites of the

modified pT73 vector. Library went through one round of

normalization."

/db_xref="taxon:9606"

/clone="IMAGE:900339"

/clone_lib="NCI-CGAP_C03"

/sex="pooled"

		/tissue_type="colon"		
		/lab_host="DH10B"		
mrna		<1.	>509	
BASE COUNT	144 a	130 c	114 g	121 t
ORIGIN				
Query Match	16.1%	Score 48;	DB 8;	Length 509;
Best Local Similarity	84.6%	Pred. No. 3.57e-46;		
Matches	55; Conservative	0; Mismatches	9; Indels	1; Gaps
Db	446 GGGGCACAGGAGGGC-AAAGATTCACCTTAAGCCAGAGATTGCGAGCTTCTGTGGTT	504		
Cp	308 GGGCCACANGNAGGGCCAAAAGTATCACTTAACCCANANNAATTGCAGCTCTGTGGTT	249		
Db	505 GGGGC	509		
Cp	248 NGGNC	244		

	8					
RESULT LOCUS DEFINITION ACCESSION NID KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS TITLE JOURNAL COMMENT	AA934377 0627g12.s1 Soares_NFL.T_GRC.S1 IMAGE:1578118 3 , mRNA sequence. AA934377 q3091534 EST.	511 bp 	mRNA 	EST 	CDNA clone	28-APR-1998
	Homo sapiens Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 511) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP). Tumor Gene Index Unpublished (1997)					

CONTACT: Robert Strausberg, Ph.D.
 Tel.: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LBNL : contact the
 IMAGE Consortium (info@image.lbnl.gov) for further information.
 Seq primer: 40m13 fwd. Et from Amersham
 High quality sequence stop: 473.
 Location/Qualifiers
 1..511
FEATURES
 source

```

/organism="Homo sapiens"
/ncbi="Organ: pooled; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI.
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBH13W, testis NHT, and B-cell
NCI.CGAP.GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682633-687230,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima BonaIido.."
/db_xref="taxon:9606"
/clone="IMAGE:1578118"
/clone.lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"

```

Query Match	16.1%	Score 48	DB 14	Length 511
Best Local Similarity	84.6%	Pred. No.	3,57e-6	
Matches	55	Conservative	0	Mismatches 9
				Indels 1
				Gaps 1
Db	448	GGGGCACAAGAGGGC	AAAAGTATCACCTTAACCCAGAGATTCGAGCTCTGTGGTT	506
		11111111111111111111	11111111111111111111	
		GGGCGACACAGNAGGGCC	AAAAGTATCACCTTAACCCAGAGATTCGAGCTCTGTGGTT	249

Db	507	GGGGC	511
Cp	248	NGGNC	244

RESULT	9	AA754459	252 bp	mRNA	EST	20-JAN-1998
LOCUS		97SN1787				
DEFINITION		Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa				
		cDNA clone 97SN1787, mRNA sequence.				
ACCESSION		AA754459				
NID		92801165				
KEYWORDS		EST.				
SOURCE		rice.				
ORGANISM		Oryza sativa				

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
<i>Oryza sativa</i>	Embryophyta; Viridiplantae; Charophyta/Embryophyta group; Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.	Nahm, B. H., Kim, J. K., Cheong, J. J., Kim, S. I., Hahn, T. R., Moon, E. P., Kim, W. T., Kim, W. Y., Yang, M. S., Park, R. D., Sohn, U. I., Kang, K. Y., Lee, M. C. and Eun, M. Y.	Large-scale Sequencing Analysis of ESTs from Rice Immature Seed	Unpublished (1998)	

CONTACT: Eun M.Y.
 Department of Cytogenetics
 National Inst. of Agri. Sci. and Tech, RDA
 Suwon, Kyunggi-do, Korea
 Tel.: 82 331 290 0301
 Fax: 82 331 290 0307
 Email: myeun@sun20.asti.re.kr
 Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
 University, Yongin, Korea. 449-728 bhnahe@naver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
 Location/Qualifiers
 1..252

```

/organism="Oryza sativa"
/cultivar="Milyang23"
/notes="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAP11
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/clone="97SN1787"
/clone_lib="Rice Immature Seed lambda ZAP11 cDNA library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"

```

	Query Match	15.8%	Score 47	DB 12	Length 252
	Best Local	Similarity 11.1%	Pred. No. 2.21e-44		
	Matches 21	Conservative	95	Mismatches 69	Indels 4
				Gaps 4	
Db	36	SHGNYMSVHNCBTRBGTHCDCKKNVNMSTMTWGVNMBVNSGDHMYBWBNTKYDVGHHTRC	95		
QY	125	CAAAINTGCGAGATCTCTACTGACATCTCTGGGACACTGGCCACAGNGTGGCTTCATAC	184		
Db	96	SRMBVTRMAHYHDTNCBRYNNNDYMMHBMBYBTHGCTCTMMGMBHNTRKCTASGMH	155		
QY	185	ACTGTACCCACGCTCTTTCTTAAAGAGAAAGTGTGGGACACCTTGACACTTTTGGGTG	243		
Db	156	TSTNYDKSSNTNTGVTBSYDKSMHGWCBSBKYHTKSTYTRATRSYTCVRKRYMMWT	215		
QY	244	-GNCNNAACAC-AGGANCTGCAATNTNTGNTTAGGT-GATACTTTGGCCCTCN	300		
Db	216	KKKYKKYHY	224		
QY	301	TGTGGCCCC	309		

RESULT	10			
LOCUS	AQ065383	344 bp	DNA	GSS
				03-AUG-1998

DEFINITION HS.2224.B1.C10.MR.CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2224 Col=19 Row=F, genomic survey sequence.

ACCESSION A065383

NID 93380901

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota: Metazoa; Chordata; Vertebrata: Mammalia: Eutheria; Primates: Catarrhini: Hominoidea: Homo.

AUTHORS 1 (bases 1 to 344)

TITLE Mahtiras,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S., Trischoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E. Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors

JOURNAL Unpublished (1997)

COMMENT

Contact: Mahtiras GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2224 Row: F Column: 19
Class: BAC ends
High quality sequence stop: 344.

FEATURES

Location/Qualifiers

1..344

/organism="Homo sapiens"

/note="Organ: Sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

/db_xref="taxon:9606"

/clone="Plate=2224 Col=19 Row=F"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

BASE COUNT 95 a 80 c 82 g 87 t

ORIGIN

Query Match 14.1%; Score 42; DB 27; Length 344;
Best Local Similarity 73.9%; Pred. No. 1.38e-35;
Matches 68; Conservative 0; Mismatches 23; Indels 1; Gaps 1;

Db 47 TATCAAGCCAAATGTACAAATGCCAATG-ACACTCTGGATATTGAGCAGGAGGCTGC 105
|||||
Oy 117 TATCAAGCCAAATGTACAAATGCCAATG-ACACTCTGGATATTGAGCAGGAGGCTGC 176
|||||
Db 106 CCTCGTACATTGTACCGTCTGCTCCCTTGAA 137
|||||
Oy 177 CTTCATACACTGTACCCAGCTCTTTCTTAA 208
|||||

RESULT 11

LOCUS AA754458 247 bp mRNA EST 20-JAN-1998

DEFINITION 97SN1784 Rice Immature Seed Lambda ZAP11 cDNA Library Oryza sativa

ACCESSION AA754458

NID 92801164

KEYWORDS EST.

SOURCE rice.

ORGANISM Oryza sativa

REFERENCE Eukaryota: Viridiplantae; Charophyta/Embryophyta group; Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.

AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.

TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed

JOURNAL Unpublished (1998)

COMMENT

Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agril. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.astl.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bnhahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse primer.

FEATURES

Location/Qualifiers

1..247

/organism="Oryza sativa"

/cultivar="Milyang25"

/note="Vector: Bluescript SK(+); Site_1: EcoRI; Site_2: XhoI; Directional cDNA library inserted into lambda ZAP11 vector at 5' end with EcoRI and 3' end with Xho I site."

/db_xref="taxon:4530"

/clone="97SN1784"

/clone_lib="Rice Immature Seed Lambda ZAP11 cDNA Library"

/tissue_type="Immature Seed"

/dev_stage="5 days after pollination"

/lab_host="E. coli SOLR"

BASE COUNT 7 a 16 c 21 g 34 t 169 others

ORIGIN

Query Match 13.8%; Score 41; DB 12; Length 247;
Best Local Similarity 11.8%; Pred. No. 7.33e-34;
Matches 22; Conservative 87; Mismatches 76; Indels 2; Gaps 2;

Db 30 BMCVRRYGTTTNGKNHGRITTTNDCSDNAHCRTYTBWYBARKYGGYGBYISMNVDN 89
|||
Oy 80 CTCTAAGCCACAGAGATTTTATTAATGTATCAACCCAAATNTGAGATGC 139
|||
Db 90 TGTGVCRTYTVNHSWNRC-SNSVYVMBTAYCDYBHDNRHVDTRCTDRGY-C 147
|||
Oy 140 TACTGCGATCTTGGGAACTGGGACACAGAGNGTGGCTTCATCATGACATGCCACTC 199
|||
Db 148 NYTASDNGTSATKRYTGYKTDSDCGGCMWRKYTYSBRYBCGVNMYVTTSMPTDKST 207
|||
Oy 200 TTTTCTAAGAAGAGATNGTGGCAGACTGTGCTGTCGCCNACACAGAGAN 259
|||
Db 208 KMSMDM 214
|||
Oy 260 GCTGCAA 266
|||

RESULT 12

LOCUS AA754458 247 bp mRNA EST 20-JAN-1998

DEFINITION 97SN1784 Rice Immature Seed Lambda ZAP11 cDNA Library Oryza sativa

ACCESSION AA754458

NID 92801164

KEYWORDS EST.

SOURCE rice.

ORGANISM Oryza sativa

REFERENCE Eukaryota: Viridiplantae; Charophyta/Embryophyta group; Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.

AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.

TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed

JOURNAL Unpublished (1998)

COMMENT

Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agril. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.astl.re.kr

QY 97 TTTTATAATGTTAATGTTATCAAGCAATATGCGATGCT 140

RESULT 15
LOCUS AA823012 337 bp mRNA EST 17-FEB-1998
DEFINITION vw39f03.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
1246205.5', mRNA sequence.
ACCESSION AA823012
NID 92892880
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryotae: Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 337)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT

Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:659893
Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 299.
Location/Qualifiers

FEATURES

source

1..337
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Organ: mammary gland; Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
TGTACCAATCTGACGAGCGGCGCCGCAATGTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
/db_xref="taxon:10090"
/clone="1246205"
/clone_1lb="Soares mouse mammary gland NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"

BASE COUNT 77 a 87 c 99 g 74 t
ORIGIN

Query Match 10.4%; Score 31; DB 13; Length 337;
Best Local Similarity 84.1%; Pred.No.1.85e-17;
Matches 37; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 119 TTTTCATGATGTATTAATATATCAGCCAAAGCTGCAGATGCT 162
|||||
QY 97 TTTTATAATGTTAATGTTATCAAGCAATATGCGATGCT 140

Search completed: Tue Dec 22 05:50:16 1998
Job time : 473 secs.


```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HL60"
/cell_type="promyelocyte"
/dev_stage="adult"
/sex="female"

BASE COUNT      102 a      42 c      38 g      114 t
ORIGIN

Query Match      100.0%; Score 296; DB 7; Length 296;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GATCGCCCATAGAAATCTAGTCACTTAATTTAGTAGTAATTAATTCGAGGT 60
    |||||||
Qy 1 GATCGCCCATAGAAATCTAGTCACTTAATTTAGTAGTAATTAATTCGAGGT 60
    |||||||

Db 61 AATTGTTTTACAAAGAACCACTGACTCCCTAATGCATTAAATATTTTATTTA 120
    |||||||
Qy 61 AATTGTTTTACAAAGAACCACTGACTCCCTAATGCATTAAATATTTTATTTA 120
    |||||||

Db 121 AATACTTTATTTATACCTTTAGAAAACATGATGTTGTTAAACATCTGTTCTTC 180
    |||||||
Qy 121 AATACTTTATTTATACCTTTAGAAAACATGATGTTGTTAAACATCTGTTCTTC 180
    |||||||

Db 181 AGTATTTTCATTTGGAACTCCATAGGGCAATGATGAAGTATATCTCTCT 240
    |||||||
Qy 181 AGTATTTTCATTTGGAACTCCATAGGGCAATGATGAAGTATATCTCTCT 240
    |||||||

Db 241 TGTAGTACATGATGATCCACAGACACTCAATAACTTTTGGTGTAACTGAAA 296
    |||||||
Qy 241 TGTAGTACATGATGATCCACAGACACTCAATAACTTTTGGTGTAACTGAAA 296
    |||||||
```

```
RESULT 2
LOCUS AA906845 413 bp mRNA EST 10-JUN-1998
DEFINITION o120a09.s1 Soares_NFL.T.GBC.S1 Homo sapiens cDNA clone
IMAGE:1523992 3', mRNA sequence.
ACCESSION AA906845
NID 93042089
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 413)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1095 Std Error: 0.00
Seq primer: -40ml3 fwd. Et from Amersham
High quality sequence stop: 351.
Location/Qualifiers
1. 413
/organism="Homo sapiens"
/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung Nhlh19W, testis NHT, and B-cell
NCI-CGAP-GCB1) were mixed and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
```

```
Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/cclone="IMAGE:1523992"
/cclone_id="Soares_NFL.T.GBC.S1"
/lab_host="DH10B"

BASE COUNT      155 a      53 c      54 g      151 t
ORIGIN

Query Match      100.0%; Score 296; DB 15; Length 413;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 TTTCAGTTTAAACCAAAAGTTATAGAGTCTGTGGATCATTTGTTCTCAAGAG 66
    |||||||
Cp 236 TTTCAGTTTAAACCAAAAGTTATAGAGTCTGTGGATCATTTGTTCTCAAGAG 237
    |||||||

Db 67 ACAGATTAATTAATCTTCATTCATTTGGCCATTTGGACCTTCCAATGAATAATCTGAAG 126
    |||||||
Cp 236 ACAGATTAATTAATCTTCATTCATTTGGCCATTTGGACCTTCCAATGAATAATCTGAAG 177
    |||||||

Db 127 AACAAATGATGTTTAAACATATCATGATGTTCTTAAAGTTTAAATTAATTTTAA 186
    |||||||
Cp 176 AACAAATGATGTTTAAACATATCATGATGTTCTTAAAGTTTAAATTAATTTTAA 117
    |||||||

Db 187 TAAATATTTTATGATGATGAGGAGTCAAGTGGATTTCTTGTAAACATTTACT 246
    |||||||
Cp 116 TAAATATTTTATGATGATGAGGAGTCAAGTGGATTTCTTGTAAACATTTACT 57
    |||||||

Db 247 GCCATTTATTTACTACATTAATAATTTAGAGTGAACATGATTTCTTATGCGAGATC 302
    |||||||
Cp 56 GCCATTTATTTACTACATTAATAATTTAGAGTGAACATGATTTCTTATGCGAGATC 1
    |||||||
```

```
RESULT 3
LOCUS AA836401 447 bp mRNA EST 31-MAR-1998
DEFINITION o637d12.s1 NCI-CGAP-GCB1 Homo sapiens cDNA clone IMAGE:1370135,
mRNA sequence.
ACCESSION AA836401
NID 92910720
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 447)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bdrip/image/image.html
Insert Length: 960 Std Error: 0.00
Seq primer: -40ml3 fwd. Et from Amersham
High quality sequence stop: 418.
Location/Qualifiers
1. 447
/organism="Homo sapiens"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD+),
```

provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTACCAATCTGAAGTGGAGCGCGCCCTCATTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo.

/db_xref="taxon:9606"

/clone_1ib="NCI_CGAP_GCB1"

/tissue_type="germinal center B cell"

/lab_host="DH10B"

BASE COUNT 164 a 55 c 56 g 172 t
ORIGIN
Query Match 100.0%; Score 296; DB 14; Length 447;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 23 TTTCAGTTAACACCAAAAGTTATTGAGTGTCTGTGATACATGTGTACTACAGAG 82
|||||
Cp 296 TTTCAGTTAACACCAAAAGTTATTGAGTGTCTGTGATACATGTGTACTACAGAG 237
Db 83 ACAGATTAATTAATCTCATTCATTTGCCCTATTGACCTCCAAATGAAAAATCTGAAG 142
|||||
Cp 236 ACAGATTAATTAATCTCATTCATTTGCCCTATTGACCTCCAAATGAAAAATCTGAAG 177
Db 143 AACAAATGATGTTTAAACAAATCTCATTCCTTAAAGTTTAAATTAAGTTATTAA 202
|||||
Cp 176 AACAAATGATGTTTAAACAAATCTCATTCCTTAAAGTTTAAATTAAGTTATTAA 117
Db 203 TAAATTAATTTTAAATGATTAAGGGAAGTCAAGTGTGATTTCTTTTAAACAAATTA 262
|||||
Cp 116 TAAATTAATTTTAAATGATTAAGGGAAGTCAAGTGTGATTTCTTTTAAACAAATTA 57
Db 263 GCCATTTATTACTACATAAATTAAGTGAATTAAGTGAATTTCTTTTAAAGCAATC 318
|||||
Cp 56 GCCATTTATTACTACATAAATTAAGTGAATTAAGTGAATTTCTTTTAAAGCAATC 1

RESULT 4 AA748329 342 bp mRNA EST 23-JAN-1998
LOCUS 0a54e10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:130810,
DEFINITION mRNA sequence.
ACCESSION AA748329
NID 92788287
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 342)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/birp/image/image.html
Insert Length: 1525 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 321.
Location/Qualifiers

FEATURES
source

1. 342

/organism="Homo sapiens"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTACCAATCTGAAGTGGAGCGCGCCCTCATTTTTTTTTTTTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_1ib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"

BASE COUNT 128 a 42 c 46 g 126 t
ORIGIN
Query Match 98.0%; Score 290; DB 12; Length 342;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 8 TTACACACCAAAAGTTATTGAGTGTCTGTGATACATGTGTACTACAGACAGAT 67
|||||
Cp 290 TTACACACCAAAAGTTATTGAGTGTCTGTGATACATGTGTACTACAGACAGAT 231
Db 68 AATAATCTCATTCATTCATTTGCCCTATTGACCTCCAAATGAAAAATCTGAAGACAA 127
|||||
Cp 230 AATAATCTCATTCATTCATTTGCCCTATTGACCTCCAAATGAAAAATCTGAAGACAA 171
Db 128 TGATGTTTAAACAAATCTCATTCCTTAAAGTTTAAATTAAGTTATTAAAT 187
|||||
Cp 170 TGATGTTTAAACAAATCTCATTCCTTAAAGTTTAAATTAAGTTATTAAAT 111
Db 168 TATTTTAAATGATTAAGGGAAGTCAAGTGTGATTTCTTTTAAACAAATTA 247
|||||
Cp 110 TATTTTAAATGATTAAGGGAAGTCAAGTGTGATTTCTTTTAAACAAATTA 51
Db 248 TATTTTAAATGATTAAGGGAAGTCAAGTGTGATTTCTTTTAAAGCAATC 297
|||||
Cp 50 TATTTTAAATGATTAAGGGAAGTCAAGTGTGATTTCTTTTAAAGCAATC 1

RESULT 5 AA808708 403 bp mRNA EST 12-FEB-1998
LOCUS 0a47d05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1308105,
DEFINITION mRNA sequence.
ACCESSION AA808708
NID 92878114
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 403)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.

University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104

Tel: 215-898-9384
Fax: 215-898-8780

Email: jecker@genome.bio.upenn.edu

Seq primer: Sp6

Class: BAC ends.

FEATURES

source

Location/Qualifiers

1. .338

/organism="Arabidopsis thaliana"

/strain="Columbia"

/note="Vector: BelosACII; Site_1: HindIII; Site_2:

HindIII; Produced by Rod Wing"

/db_xref="taxon:3702"

/clone_id="TAMU"

/sex="hermaphrodite"

BASE COUNT 117 a 0 c 61 g 160 t

ORIGIN

Query Match

Best local Similarity 71.2%; Pred. No. 4.71e-05;

Matches 47: Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Db 1 TTGCAATTAAAGATTATTATTAATATGTTAATTATTATTAAGATTAA 60

111 TTTTATTAATTAACCTTATTATTAACCTTGAACATGATGTTTAAACATCA 170

Db 61 TTGT 65

171 TTGT 176

171 TTGT 176

RESULT 13 AA089382 235 bp mRNA EST 12-DEC-1996

LOCUS SM3ICA1587SK Brugia malayi infective larva cDNA (SAM94WL-Bml3)

DEFINITION Brugia malayi cDNA clone SM3ICA1587 5', mRNA sequence.

ACCESSION AA089382

NID 91634861

KEYWORDS EST.

SOURCE Brugia malayi.

ORGANISM Brugia malayi.

Eukaryota; Eukaryotes; Metazoa; Nematoda;

Secernentea; Spirurida; Spirurina; Filarioidea;

Onchocercidae; Brugia.

1 (bases 1 to 235)

REFERENCE 1 Williams, S. A. Genes expressed in infective third stage larvae of Brugia malayi

UNPUBLISHED (1995)

COMMENT Contact: Steven A. Williams

Molecular Parasitology

Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA

Tel: 4135853826

Fax: 4135853786

Email: genome@smith.edu

Seq primer: pBluescript SK.

Location/Qualifiers

1. .235

/organism="Brugia malayi"

/strain="TFS Labs"

/note="Vector: lambda Unizap XR; Site_1: EcoR I; Site_2:

Xho I; Lymphatic filarial nematode parasite of humans;

Brugia malayi isolated from mosquitoes 10 days after

infection and converted to double stranded cDNA using

reverse transcriptase and oligo(dT) followed by RNase H

and DNase I. The library had 1.6 x 10⁶ independent

recombinants and average insert size was 900 base pairs.

The library was constructed by Wennong Lu. The library is

available from Dr. S. A. Williams, email genome@smith.edu."

/db_xref="taxon:6279"

/clone="SM3ICA1587"

/clone_id="Brugia malayi infective larva cDNA

(SAM94WL-Bml3)"

/lab_host="XLI-Blue MRF"

<1. .>235

BASE COUNT 70 a 9 c 19 g 134 t 3 others

ORIGIN

Query Match 9.1%; Score 27; DB 22; Length 235;

Best local Similarity 64.0%; Pred. No. 3.25e-04;

Matches 57: Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Db 116 ATTCTGTTTTTTTATATTAAGTAANTTAATTTNCACTTTATATAGTTTAA 175

108 ATATTTTATTAATTAACCTTATTATTAACCTTAGAACAATGATGTTTAA 167

Db 176 TATTTTATTAATTTGTTTATTTGG 204

168 TCATTTGTTCTTCAGTATTTTCATTTGG 196

RESULT 14 N79132 215 bp mRNA EST 02-APR-1996

LOCUS y292d10.r1 Homo sapiens cDNA clone 290515 5'.

DEFINITION N79132

ACCESSION N79132

NID 91241833

KEYWORDS EST.

SOURCE human clone-290515 primer-reverse ET library-Soares multiple

sclerosis 2NDMSD vector-pT73D (Pharmacia) with a modified

polylinker V.LTYPE: Phagemid host-DH10B (ampicillin resistant)

Rsite1=Not I Rsite2=Eco RI 46 year old male. 1st strand cDNA was

primed with a Not I - oligo(dT) primer

[5'-TGTTACCAATCGAAGTGGAGCGCGCATTTTCTTTTCTTTT-3']

double-stranded cDNA was size selected, ligated to Eco RI adapters

(Pharmacia), digested with Not I and cloned into the Not I and Eco

RI sites of a modified pT7 vector (Pharmacia). Library went

through one round of normalization to a Cot = 5. Library

constructed by Bento Soares and M.Fatima Bonaldo. RNA from 4

multiple sclerosis lesions from one patient was kindly provided by

Dr. Kevin G. Becker (NINDS/NIH).

Homo sapiens

Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata;

Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;

Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;

Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 215)

REFERENCE 1 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

Trevasakis, E., Waterston, R., Williamson, A., Woldmann, P. and

Wilson, R.

COMMENT The WashU-Merck EST project

UNPUBLISHED (1995)

TITLE Wilson RK

JOURNAL WashU-Merck EST Project

COMMENT Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

High quality sequence stops: 184

Source: IMAGE Consortium, LLNL.

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Location/Qualifiers

1. .215

/organism="Homo sapiens"

/clone="290515"

<1. .>215

BASE COUNT 64 a 28 c 28 g 94 t 1 others

 WISE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MSPrch_n n.a. - n.a. database search, using Smith-Waterman algorithm
 Run on: Tue Dec 22 04:06:17 1998; MasPar time 59.90 Seconds
 672.347 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-530-112A-563
 Description: (1-296) from US08530112A.seq
 Perfect Score: 296
 N.A. Sequence: 1 GATCTGCCATAGAAATCT.....TTTGGTGTGTAACGTGAA 296
 Comp: CTAGACGGTATCTTTTAGA.....AAACCAACAATTTGACTTT

Scoring table: TABLE default
 Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: n-geneseq32
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39 40:part40

Statistics: Mean 7.796; Variance 6.036; scale 1.292

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description	Pred. No.
1	296	100.0	296.18	T19563 Human gene signature	2.19e-143
2	38	12.8	91.9	051746 Oligonucleotide probe	1.13e-06
3	35	11.8	204.1	N81164 Base substituted E.co	2.64e-05
4	34	11.5	204.1	N81164 Base substituted E.co	7.44e-05
5	31	10.5	91.9	051746 Oligonucleotide probe	1.59e-03
6	27	9.1	501.3	N50030 Sequence encoding new	8.28e-02
7	27	9.1	501.3	N50025 Sequence encoding new	8.28e-02
8	26	8.8	501.3	N50027 Sequence encoding new	2.17e-01
9	26	8.8	501.3	N50029 Sequence encoding new	2.17e-01
10	26	8.8	501.3	N50031 Sequence encoding new	2.17e-01
11	26	8.8	501.3	N50033 Sequence encoding new	2.17e-01
12	25	8.4	344.8	O60902 Degraderate Alteromona	5.60e-01
13	24	8.1	39.24	T14323 Human brain Expressed	1.43e+00

14	24	8.1	501.3	N50024 Sequence encoding new	1.43e+00
15	24	8.1	501.3	N50022 Sequence encoding new	1.43e+00
16	24	8.1	565.6	O35072 HCV envelope region n	1.43e+00
17	24	8.1	605.21	T31530 Human 3' Apolipoprote	1.43e+00
18	24	8.1	20303.31	T71699 Human deoxycytidylate	1.43e+00
19	24	8.1	26764.31	T71696 Human deoxycytidylate	1.43e+00
20	23	7.8	67.24	T14322 Primer used in the la	3.59e+00
21	23	7.8	68.34	T73397 Oligonucleotide tag c	3.59e+00
22	23	7.8	68.32	T63255 Messenger RNA primer	3.59e+00
23	23	7.8	70.24	T14325 Conjugate formed by 1	3.59e+00
24	23	7.8	383.8	O61360 Human brain Expressed	3.59e+00
25	23	7.8	501.3	N50028 Sequence encoding new	3.59e+00
26	23	7.8	1230.1	N90223 Malaria-specific DNA	3.59e+00
27	23	7.8	1704.1	N90725 Groundnut stilbene sy	3.59e+00
28	23	7.8	1761.17	T01407 H6-Promoted CHV gc ge	3.59e+00
29	23	7.8	3369.1	N81690 Atrial natriuretic po	3.59e+00
30	23	7.8	4590.2	N60472 Sequence encoding the	3.59e+00
31	23	7.8	4599.22	T27052 RRP3 telomerase-assoc	3.59e+00
32	23	7.8	19124.32	T72882 Plasmidium var-7 gene	3.59e+00
33	23	7.8	133894.17	T13635 AcNPV genomic DNA clo	3.59e+00
34	22	7.4	114.12	O70469 Generic DNA sequence	8.88e+00
35	22	7.4	114.12	O70465 Generic DNA sequence	8.88e+00
36	22	7.4	114.12	O70467 Generic DNA sequence	8.88e+00
37	22	7.4	501.3	N50032 Sequence encoding new	8.88e+00
38	22	7.4	501.3	N50023 Sequence encoding new	8.88e+00
39	22	7.4	1052.1	N90224 Malaria-specific DNA	8.88e+00
40	22	7.4	1979.36	T87015 Human Fe-gamma recept	8.88e+00
41	22	7.4	1990.2	N92451 CDNA Insert of vector	8.88e+00
42	22	7.4	4918.10	O56696 Sequence of the L. la	8.88e+00
43	22	7.4	5547.34	Cotton fibre promoter	8.88e+00
44	22	7.4	8342.13	O75209 ALL-1 (acute lymphocy	8.88e+00
45	22	7.4	8391.19	T16333 MLV gene 8.3 kb fragm	8.88e+00

ALIGNMENTS

RESULT 1
 ID T19563 standard; cDNA to mRNA; 296 BP.
 AC T19563;
 DT 27-JUN-1996 (first entry)
 DE Human gene signature HUNG500628.
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human, cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.
 OS Homo sapiens.
 PN WO9514772-A1.
 PD 01-JUN-1995.
 PF 11-NOV-1994; J01916.
 PR 12-NOV-1993; JF-355504.
 PA (MATS/) MATSUBARA K.
 PI (OKUBO/) OKUBO K.
 PI Matsubara K, Okubo K;
 DR WPI: 95-206931/27.
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 PS Claim 1: Page 418; 22459P; Japanese.
 CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in T19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and genes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.

```

SQ Sequence 296 BP; 102 A; 42 C; 38 G; 114 T;
Query Match 100.0%; Score 296; DB 18; Length 296;
Best Local Similarity 100.0%; Pred. No. 2.19e-143;
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 gatctgcataagaagaactagctcaactcaatttatgtatgaataaattgcaggt 60
1 gatctgcataagaagaactagctcaactcaatttatgtatgaataaattgcaggt 60
QY 1 gatctgcataagaagaactagctcaactcaatttatgtatgaataaattgcaggt 60
1 gatctgcataagaagaactagctcaactcaatttatgtatgaataaattgcaggt 60
Db 61 aattgttttcaaaagaatccaccctgactccccaatgcatataaaatattattat 120
61 aattgttttcaaaagaatccaccctgactccccaatgcatataaaatattattat 120
QY 61 aattgttttcaaaagaatccaccctgactccccaatgcatataaaatattattat 120
121 aataacttattatataactttttagaacaatgtagtattgttttaaacattgtcttc 180
121 aataacttattatataactttttagaacaatgtagtattgttttaaacattgtcttc 180
QY 121 aataacttattatataactttttagaacaatgtagtattgttttaaacattgtcttc 180
121 aataacttattatataactttttagaacaatgtagtattgttttaaacattgtcttc 180
Db 181 agtatttttcattggaagcccaataggcaattgaatgaatattatctgtctct 240
181 agtatttttcattggaagcccaataggcaattgaatgaatattatctgtctct 240
QY 181 agtatttttcattggaagcccaataggcaattgaatgaatattatctgtctct 240
181 agtatttttcattggaagcccaataggcaattgaatgaatattatctgtctct 240
Db 241 tctagtaacaatgtagtccacagacacactcaataaactttgtgtttaactgaa 296
241 tctagtaacaatgtagtccacagacacactcaataaactttgtgtttaactgaa 296
QY 241 tctagtaacaatgtagtccacagacacactcaataaactttgtgtttaactgaa 296
241 tctagtaacaatgtagtccacagacacactcaataaactttgtgtttaactgaa 296

RESULT 2
ID 051746 standard; cDNA; 91 BP.
AC 051746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KM Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PE 24-MAY-1993: 108325.
PR 26-MAY-1992: US-889651.
PA (BECT ) BECTON DICKINSON CO.
PI Shank DD; Spears PA;
DR WPI; 93-378844/48.
PT New oligo: nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (051735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also 051735-45 and 051747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 12.8%; Score 38; DB 9; Length 91;
Best Local Similarity 0.0%; Pred. No. 1.13e-06;
Matches 0; Conservative 43; Mismatches 5; Indels 0; Gaps 0;

Db 13 vhsyvvvvhvshhshvhhvhhvhhvhhvhhvhhvhhvhhvhhvhhvhhvhhv 60
13 vhsyvvvvhvshhshvhhhshvhhvhhvhhvhhvhhvhhvhhvhhvhhvhhvhhv 60
Cp 262 CTGTTGATACATGTACTACACAGACAGATATATATCTTCAATTC 215

RESULT 3
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KM E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT /tag- a
FT /function-multiple cloning site

```

```

FT primer_bind 187..204
FT /tag- b
PN EP-285123-A.
PD 05-MAY-1988.
PE 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR WPI; 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; P; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 11.8%; Score 35; DB 1; Length 204;
Best Local Similarity 9.1%; Pred. No. 2.64e-05;
Matches 9; Conservative 51; Mismatches 39; Indels 0; Gaps 0;

Db 87 mtthhyrmhbydyndrsdaayccyrsvkydcocnchdhvbybbvynv 146
127 agtatttttaataaataaatttttaattgacatgaggaagcagctgattctgtgaaa 68
Cp 127 agtatttttaataaataaatttttaattgacatgaggaagcagctgattctgtgaaa 68
147 nhnnccebnhvhvhhvhhvhhvhhvhhvhhvhhvhhvhhvhhvhhvhhvhhvhhv 185
67 AACATTACCGCCCAATTATTACTACATAAATTTGA 29

RESULT 4
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KM E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT /tag- a
FT /function-multiple cloning site
FT primer_bind 187..204
FT /tag- b
PN EP-285123-A.
PD 05-MAY-1988.
PE 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR WPI; 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; P; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.

```

CC The sequence covers all 176 diffit base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575
SO Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match
Best Local Similarity 11.5%; Score 34; DB 1; Length 204;
Matches 10; Conservative 64; Mismatches 42; Indels 2; Gaps 2;

Db 75 yycdchvgccgymrttthyrtrmbvryrdynrsd-aaawyc-yrrsvkydcynach 132
Oy 175 TTTCTTCAGTATTTTCATTGAGAGTCACATAGGCCAAATTCAGATATATATATCT 234

Db 133 ddhybbvbbvynhnmccbnhvhbhmhnmhwayrharrrdvhccvch 190
Oy 235 GTCCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 292

RESULT 5
ID 051746 standard; cDNA; 91 BP.
AC 051746: (first entry)
DT 31-MAY-1994
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PR 24-MAY-1993; 108325.
PS 26-MAY-1992; US-889651.
PA (BECT) BECTON DICKINSON CO.
PI Shank DD, Spears PA.
DR WPI: 93-37884/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3: Page 14; 23pp: English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match
Best Local Similarity 10.5%; Score 31; DB 9; Length 91;
Matches 0; Conservative 37; Mismatches 6; Indels 0; Gaps 0;

Db 13 vhsyvvvshvshvshvshvshvshvshvshvshvshvshvshvshvshvshvshvsh 55
Oy 237 CTTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 279

RESULT 6
ID N50030 standard; DNA; 501 BP.
AC N50030: (first entry)
DT 04-SEP-1991
DE Sequence encoding new modified human beta interferon polypeptides
DE IFN 447.
KW Antiviral; cell growth regulator; immune system regulator;
KW antiproliferative; ss.
OS Homo sapiens.
FH Key
FT cds
FT Location/Qualifiers
FT 1..501
FT /*tag= a

PN EP-163993-A.
PD 11-DEC-1985.
PR 17-MAY-1985; 105750.
PS 17-MAY-1984; GB-012564.
PA (SEAR) SEARLE G D & CO.
PI Bell ID, Boseley PG, Porter AG;
DR WPI: 85-311944/50.
P-PSDB: P50029.

PT New modified human beta interferon polypeptide(s) - prepd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PS Claim 28; Chart 2h, page 39; 71pp: English.
CC Compared with interferon beta prepd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index; improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixes.
SQ Sequence 501 BP; 110 A; 31 C; 69 G; 79 T;

Query Match
Best Local Similarity 9.1%; Score 27; DB 3; Length 501;
Matches 38; Conservative 43; Mismatches 43; Indels 3; Gaps 3;

Db 10 aaytybtdgnttytbtbarmgdmnnwnaay-tyccarfyccaraaytybtdgca 68
Oy 77 AATCCACCTGACTTCCCTCATGATGATTAATAATTTTATTTAATACTTATTATTA 136

Db 69 rylbaaygmngdytbgartaytytbaargaymgdcagaytytymgathccncarga 128
Oy 137 ACTTTAGAAACATGT-AGTATTTGTTT-AAACATCATTTGTTCTTCAGTATTTTCATT 194

Db 129 rgarthy 135
Oy 195 GGAAATC 201

RESULT 7
ID N50025 standard; DNA; 501 BP.
AC N50025: (first entry)
DT 04-SEP-1991
DE Sequence encoding new modified human beta interferon polypeptides
DE IFN 418.
KW Antiviral; cell growth regulator; immune system regulator;
KW antiproliferative; ss.
OS Homo sapiens.
FH Key
FT cds
FT Location/Qualifiers
FT 1..501
FT /*tag= a

PN EP-163993-A.
PD 11-DEC-1985.
PR 17-MAY-1985; 105750.
PS 17-MAY-1984; GB-012564.
PA (SEAR) SEARLE G D & CO.
PI Bell ID, Boseley PG, Porter AG;
DR WPI: 85-311944/50.
P-PSDB: P50024.

PT New modified human beta interferon polypeptide(s) - prepd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PS Claim 28; Chart 2c, page 34; 71pp: English.
CC Compared with interferon beta prepd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index; improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixes.
SQ Sequence 501 BP; 112 A; 30 C; 69 G; 85 T;

Query Match
Best Local Similarity 9.1%; Score 27; DB 3; Length 501;
Matches 40; Conservative 41; Mismatches 43; Indels 3; Gaps 3;

Db 10 aaytybtdgnttytbtbarmgdmnnwnaay-tyccarfyccaraaytybtdgca 68
Oy 77 AATCCACCTGACTTCCCTCATGATGATTAATAATTTTATTTAATACTTATTATTA 136

Db 69 rylbaaygmngdytbgartaytytbaargaymgdcagaytytymgathccncarga 128
Oy 137 ACTTTAGAAACATGT-AGTATTTGTTTAAAC-ATCATTTGTTCTTCAGTATTTTCATT 194


```

AC 094336;
DE 17-APR-1996 (first entry)
DE Degenere Alteromonas hanedai luciferase beta-subunit gene.
KW Luciferase; beta-subunit; diagnostic; reagent; recombinant;
OS Alteromonas hanedai.
FH Key Location/Qualifiers
FT cds 1..984
FT /tag= a
PN J07222590-A.
PD 22-AUG-1995.
PF 08-FEB-1994: 035450.
PR 08-FEB-1994: JP-035450.
PA (CHCC) CHISSO CORP.
DR WPI: 95-324955/42.
DR P-PSDB: R77379.
PT DNA encoding luciferase from fluorescent bacterium Alteromonas
PT hanedai - useful as an enzymatic label in diagnostic reagents,
PT etc.
PS Claim 2: Pages 7-8, 16pp; Japanese.
CC By transfecting a host (e.g. E. coli) with the degenerate A. hanedai
CC luciferase alpha- and beta-subunit genes 094335/36, a recombinant
CC luciferase can be produced. The luciferase can be applied as an
CC enzymatic label in various diagnostic and examination reagents
CC etc., useful in high sensitivity detection methods.
CC Note: original sequence given in non-IUPAC degenerate code.
SQ Sequence 984 BP; 224 A; 74 C; 144 G; 139 T;

Query Match 8.8%; Score 26; DB 17; Length 984;
Best Local Similarity 36.1%; Pred. No. 2,17e-01;
Matches 35; Conservative 23; Mismatches 36; Indels 1; Gaps 1;

Db 849 ycaratharylungtngaraacngngtnaaraathlyntynwntgyarws 908
Cp 196 CCAATGAAATACGTGACAGCAAA-TGATGTTTAAACATGTTCTTAAG 138
Db 909 natgaargywnaaygynaaaraayathathaa 945
Cp 137 TTATTAATAAAGTTATTTAAATAAATAATTTTAAAT 101

RESULT 12
ID 060902 standard; DNA; 344 BP.
AC 060902;
DE 16-MAR-1994 (first entry)
DE Human brain Expressed Sequence Tag EST00993.
KW Gene transcription product; genetic markers; tagging; in vivo;
KM transcription; mapping; locations; chromosomes; chromosomal; ss.
OS Homo sapiens.
PN W09316178-A.
PD 19-AUG-1993.
PR 12-FEB-1993: U01294.
PR 12-FEB-1992: US-837195.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Adams MD, Moreno RF, Venter CJ;
DR WPI: 93-372882/34.
PT Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes
PS Example 4: Page 408; 500pp; English
CC The Expressed Sequence Tag was isolated from a human brain CDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prep. of antisense sequences, probes and constructs.
CC EST00993 has a "poor" coding probability as evaluated using the
CC coding region prediction program CRM. See also 059041-061440.
SQ Sequence 344 BP; 125 A; 56 C; 62 G; 97 T;

Query Match 8.4%; Score 25; DB 8; Length 344;
Best Local Similarity 74.5%; Pred. No. 5.60e-01;

```

```

Matches 38; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Db 150 catcaaatatccatttaagaacattatcaaaacttaagagacat 200
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 100 CATTAAATAATTTTATTTAATACTTTATTTATTAACCTTTAGAAACAT 150

RESULT 13
ID T14323 standard; DNA; 39 BP.
AC T14323;
DE 16-JAN-1997 (first entry)
DE Sequence used in the labelling and sorting of nucleotide molecules.
KW Labelling; sorting; sequencing; tag; tagging; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc.feature 39
FT /tag= a
FT /mod_base= Attached to a linker molecule linked to
FT a microparticle
FT W09612039-A1.
PN 25-APR-1996.
PD 12-OCT-1995: U12678.
PR 13-OCT-1994: US-322348.
PR 19-DEC-1994: US-359295.
PA (LYNX-) LYNX THERAPEUTICS INC.
PI Brenner S;
DR WPI: 96-222023/22.
PT Labelling and sorting mols. using oligo:nucleotide tags - useful in
PT large-scale parallel operations, e.g. DNA sequencing and mRNA
PT fingerprinting
PT Disclosure; Page 20; 71pp; English.
PS Determining the nucleotide sequence (1) of a target polynucleotide
CC (1) comprises: (a) generating from T a plurality of fragments that
CC cover T; (b) attaching an oligonucleotide tag from a repertoire of
CC tags, to each fragment such that all the same fragments have the
CC same tag, and all different fragments have different tags; (c)
CC sorting the fragments by specifically hybridizing the tags with
CC their respective tag complements; (d) determining (1) of a portion
CC of each of the fragments; and (e) determining (1) of T by collating
CC the sequences of the fragments. The tagging system can be used with
CC single base sequencing methods to sequence polynucleotides up to
CC several kilobases in length. The tagging system permits many
CC thousands of fragments of a target polynucleotide to be sorted onto
CC one or more solid phase supports and sequenced simultaneously.
CC A primer which is initially used to reverse transcribe mRNA is
CC described in T14322. If the amplified product is then needed to be
CC attached to a solid phase support a sequence like this may be used.
CC The mRNA would then be removed and the second strand of cDNA
CC produced using a primer with a similar form to that described in
CC T14324. After restriction enzyme digestion, the conjugate would
CC have the formula described in T14325.
SQ Sequence 39 BP; 0 A; 0 C; 11 G; 1 T;

Query Match 8.1%; Score 24; DB 24; Length 39;
Best Local Similarity 8.3%; Pred. No. 1.43e+00;
Matches 3; Conservative 27; Mismatches 6; Indels 0; Gaps 0;

Db 2 wwwgwmgwmgwmgwmgwmgwmgwmgwmgwmgwmgwmgwmgwmgwmgw 37
Cp 141 AAAGTTAATAATAAGTTATTAATAATAATAATAATTT 106

RESULT 14
ID N50024 standard; DNA; 501 BP.
AC N50024;
DE 04-SEP-1991 (first entry)
DE Sequence encoding new modified human beta interferon polypeptides
DE IFNk 417.
DE Antiviral; cell growth regulator; immune system regulator;
KW antiproliferative; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 1..501

```

```

FT      EP-163993-A.                /*tag= a
PN      11-DEC-1985.
PD      17-MAY-1985; 105750.
PR      17-MAY-1984; GB-012564.
PA      (SEAR ) SEARLE G D & CO.
PI      Bell LD, Boseley PG, Porter AG;
DR      WPI: 85-311944/50.
P-PSDB: P50023.
PT      New modified human beta interferon polypeptide(s) - prep. by
PT      plasmid transformed bacteria, with improved antiviral,
PT      anti-proliferative and immune regulating actions
PS      Claim 28: Chart 2b, page 33; 71pp: English.
CC      Compared with interferon beta prep. by recombinant methods, the
CC      INFs of the invention are more active and have different affinities
CC      for cell surface receptors (allowing selective targeting); they
CC      have higher therapeutic index; improved stability against microbial
CC      breakdown during synthesis; and better in vivo solubility and
CC      stability. They are also easier to recover from incubation mixts.
SO      Sequence 501 Bp; 110 A; 32 C; 66 G; 81 T;

```

Query Match

Best Local Similarity 28.1%; Score 24; DB 3; Length 501;

Matches 26; Conservative 32; Mismatches 30; Indels 2; Gaps 2;

```

Db      10 aaytlytbgnttytbcarmgdmnmnaay-tycarwncaraarytlytbgca 68
Oy      77 AATCCACCTGACTTCCCTCATGATTAATAATTTTATTAAATACATTATTATA 136
Db      69 rybbaargmgdytbgartaytgyybaa 98
Oy      137 ACTTTTAGAACAATGT-AGTATTGTTTAA 165

```

RESULT 15

ID N50024 standard; DNA; 501 Bp.

AC N50024:

DT 04-SEP-1991 (first entry)

DE Sequence encoding new modified human beta interferon polypeptides

DE IFN γ 417.

KW Antiviral; cell growth regulator; immune system regulator;

KW antiproliferative; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT cds 1..501

FT EP-163993-A. /*tag= a

PN 11-DEC-1985.

PD 17-MAY-1985; 105750.

PR 17-MAY-1984; GB-012564.

PA (SEAR) SEARLE G D & CO.

PI Bell LD, Boseley PG, Porter AG;

DR WPI: 85-311944/50.

DR P-PSDB: P50023.

PT New modified human beta interferon polypeptide(s) - prep. by

PT plasmid transformed bacteria, with improved antiviral,

PT anti-proliferative and immune regulating actions

PS Claim 28: Chart 2b, page 33; 71pp: English.

CC Compared with interferon beta prep. by recombinant methods, the

CC INFs of the invention are more active and have different affinities

CC for cell surface receptors (allowing selective targeting); they

CC have higher therapeutic index; improved stability against microbial

CC breakdown during synthesis; and better in vivo solubility and

CC stability. They are also easier to recover from incubation mixts.

SO Sequence 501 Bp; 110 A; 32 C; 66 G; 81 T;

Query Match 8.1%; Score 24; DB 3; Length 501;

Best Local Similarity 28.8%; Pred. No. 1.43e+00;

Matches 23; Conservative 25; Mismatches 32; Indels 0; Gaps 0;

```

Db      20 gnttytbcarmgdmnmnaaytlycarwncaraarytlytbgcarytbaargnm 79
Cp      148 GTTTCTAAAGTTATTAATTAAGTTATTAAATTAATAATTTTAAATGATTAAGGGGAA 89

```

```

Db      80 gdytbgartaytgyybaa 99
Cp      88 GTCAGTGTGATTCCTTGTAA 69

```

Search completed: Tue Dec 22 04:07:22 1998
Job time : 65 secs.

(\mathbf{w}_L)

Distribution rights by Oxford Molecular Ltd

Using Smith-Waterman algorithm

MasPar time 479.29 Seconds

be

TTTTCGTTGTTAAACTGAAA 296

A...AAAACCAACAATTGACTTT

_htg 18:gb_in 19:gb_om 20:gb_ov
 _p1 24:gb_p12 25:gb_pr1
 _ro 29:gb_st 30:gb_sts 31:gb_sy

51; scale 1.366

dicted by chance to have a score of the result being printed score distribution.

Description	Pred. No
<i>Drosophila melanogaster</i>	6, 17e+0
<i>Drosophila melanogaster</i>	5, 50e+0
<i>Anguilla anguilla</i> mito	2, 84e+0
<i>C.eichmanni</i> mitochond	7, 32e+0
Sequence 14 from paten	7, 32e+0
<i>Caenorhabditis elegans</i>	7, 32e+0
Homo sapiens Chromosom	7, 32e+0
<i>Anguilla anguilla</i> mito	1, 86e+0
<i>Anguilla anguilla</i> mito	1, 86e+0
<i>Anguilla anguilla</i> mito	1, 86e+0
<i>Anguilla anguilla</i> mito	1, 86e+0
<i>Anguilla anguilla</i> mito	1, 86e+0
<i>Elencus japonica</i> 18S	1, 86e+0

15	26	8.8	19374	17	AC005133	Human BAC clone RG33f	1.15e+01
16	26	8.8	144239	26	AC004015	Human DNA sequence firo	1.15e+01
17	26	8.8	133981	26	HS8211	Human DNA sequence firo	1.15e+01
18	28	9.1	4593	32	CBZ79650	C. beebei mitochondrial	4.67e+00
19	27	9.1	4798	18	CBSC04076	C. beebei mitochondrial	4.67e+00
20	27	9.1	7964	18	PCSC04015	Plasmodium falciparum	4.67e+00
21	27	9.1	33904	18	CEZK455	Cenotrichaditis elegans	4.67e+00
22	27	9.1	71803	17	AC005504	*** SEQUENCING IN PROG	4.67e+00
23	27	9.1	110858	15	HUM2634	Human PAC clone 2634 f	4.67e+00
24	27	9.1	114736	18	FEMAL3P3	Plasmodium falciparum	4.67e+00
25	27	9.1	125532	17	AC004975	*** SEQUENCING IN PROG	4.67e+00
26	27	9.1	144387	17	AC004898	*** SEQUENCING IN PROG	4.67e+00
27	27	9.1	244649	17	AC005308	*** SEQUENCING IN PROG	4.67e+00
28	27	9.1	244649	17	AC005308	*** SEQUENCING IN PROG	4.67e+00
29	27	9.1	244649	17	AC005308	*** SEQUENCING IN PROG	4.67e+00
30	26	8.8	492	20	AAJ225988	Anguilla anguilla mito	1.15e+01
31	26	8.8	493	20	AAJ225988	Anguilla anguilla mito	1.15e+01
32	26	8.8	493	20	AAJ225988	Anguilla anguilla mito	1.15e+01
33	26	8.8	493	20	AAJ225988	Anguilla anguilla mito	1.15e+01
34	26	8.8	493	20	AAJ225988	Anguilla anguilla mito	1.15e+01
35	26	8.8	2188	15	AAJ225997	Anguilla anguilla mito	1.15e+01
36	26	8.8	3402	15	BLCR042	Bacteriophage L cto, 2	1.15e+01
37	26	8.8	3727	15	FMU81184	Fusobacterium mortifer	1.15e+01
38	26	8.8	3727	15	CLOENTW110	C. botulinum gene for n	1.15e+01
39	26	8.8	23278	18	CEOSE12	Cenotrichaditis elegans	1.15e+01
40	26	8.8	36922	18	PCSC03076	Plasmodium falciparum	1.15e+01
41	26	8.8	41552	18	CEC17D12	Cenotrichaditis elegans	1.15e+01
42	26	8.8	100496	26	HS970D1	Human DNA sequence firo	1.15e+01
43	26	8.8	114736	18	FEMAL3P3	Plasmodium falciparum	1.15e+01
44	26	8.8	118153	26	HS249C1	Plasmodium falciparum	1.15e+01
45	26	8.8	133981	26	HS8211	Human DNA sequence firo	1.15e+01
46	26	8.8	144239	26	AC004015	Human BAC clone RG33f	1.15e+01
47	26	8.8	19374	17	AC005133	*** SEQUENCING IN PROG	1.15e+01

ALIGNMENTS

UT	1	DNA	INV	05-AUG-1997
TUS	AF012089	10772 bp		
ITION	Drosophila melanogaster	cysteine proteinase-1 (Cp1) gene, complete		
SSION	cds, and phenylalanyl tRNA synthetase gene, partial cds.			
AF012089				
WORDS	923052220			
RCE	fruit fly,			
GANISM	Drosophila melanogaster			
	Eukaryote; mitochondrial eukaryotes; Metazoa; Arthropoda;			
	Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;			
ERENCE	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
THONS	1 (bases 4546 to 4553)			
TTLE	Gray,J.H., Tanaka,M.M. and Sved,J.A.			
	P.-element-induced recombination in Drosophila melanogaster: hybrid			
JOURNAL	element insertion			
EDLINE	Genetics 144 (4), 1601-1610 (1996)			
ERENCE	97132596			
THONS	2 (bases 1 to 10772)			
TTLE	Gray,J.H.M., Sved,J.A., Preston,C.R. and Engels,W.R.			
	Structure of the cysteine proteinase (Cp1) gene of Drosophila			
	melanogaster and associated mutational effects			
JOURNAL	Unpublished			
ERENCE	3 (bases 1 to 10772)			
THONS	Gray,J.H.M., Sved,J.A., Preston,C.R. and Engels,W.R.			
TTLE	Direct Submission			
JOURNAL	Submitted (30-JUN-1997) School of Biological Sciences, University			
URES	of Sydney, Biology A12, Sydney University, NSW 2006, Australia			
Source	Location/Qualifiers			
	1..10772			
	/organism="Drosophila melanogaster"			
mrna	/db_xref="taxon:7227"			
	/join(872..1000,2310..2426,6476..6690,6751..7707)			
	/gene="Cp1"			
	/product="cysteine protease"			
gene	872..7707			
	/gene="Cp1"			

```

exon      872..1000
          /gene="Cp1"
          /number=1
          1001..2309
          /gene="Cp1"
          2310..2426
          /gene="Cp1"
          /number=2
          join(2328..2426,6476..6690,6751..7462)
          /gene="Cp1"
          /codon_start=1
          /product="cysteine proteinase-1"
          /db_xref="PID:g2305221"
          /translation="MRTAVLILPLIALLAVAQAVSFADVWMEWHTFKLEHRKNYODT
          EERFKRIFENKHKIAKHNQRPAREKVSFKLAVNKADLIHHEFRLANGFNITLHK
          QLRADSEKGVTFISPAHYTLPSVDMRKGAVTAVKDGHCSCWAFSSSTALEGO
          HFRKSGVLSLSQNLVDCSTKYGNCGNLMDNARYIKDNGIDITEKSEYEAID
          DSCHPFKGVATDRCGFTDIPQDGEKMAAVATGVSVAIDASHESFPQYSEGVAN
          EPCQDAONLDHGVLVGFTGDESGEDYWLKNSMGTWGDGKFIKMLRNKNCGLIAS
          ASSYPLV"
          2427..6475
          /gene="Cp1"
          4546..4553
          /gene="Cp1"
          /note="insertion site of P(casper)(50C)"
          /citation=[1]
          6476..6690
          /gene="Cp1"
          /number=3
          6691..6750
          /gene="Cp1"
          6751..7707
          /gene="Cp1"
          /number=4
          join(<8110..9300,9370..>9532)
          /product="phenylalanyl tRNA synthetase"
          join(8110..9300,9370..>9532)
          /note="potential orf"
          /codon_start=1
          /product="phenylalanyl tRNA synthetase"
          /db_xref="PID:g2305222"
          /translation="MLTLRVGARGAHLKSTRCLASSAPAKSPSPPOLEAGSTVA
          TDCMTNVTPEKILSYVGANKHLOTDHPILSITRORVANYFAGVAGRNCPLEFSYDGN
          PVTVQONFENPLILPADHVSROKSDCYIINOQHILRHHTAHQVELISGLDNFLVYG
          EYVRDEIDSTHYVEFHQADAVRLVTKDKLFEKRPGLLEFEETWSGTLAPKLLPBP
          SSWTKNSPAPTRARVAKLMEHEMKHVLGKLDLFGPRIKYRWDTYFPFQPSWELBI
          YFQDNMLEVIGCCIMRHEILQRSVHOSIGAVGGLERLAVLFDIPDIRLWMSNS
          GFLOSQSEKDLNHLPKYKPISHYPOCTNDSEFWLPDIEVDAGSPNDEYDLYRSVAG
          DVEQISLVQKFKHPTKGSVCFRIRYIRHMERLTLQAEVNEIHKQIASASVDSFVQ
          IX"

BASE COUNT      2929 a      2357 c      2282 g      3046 t      158 others
ORIGIN
Query Match      15.9%; score 47; DB 18; length 10772;
Best Local Similarity 22.4%; Pred. No. 6,17e-09;
Matches 34; Conservative 68; Mismatches 49; Indels 1; Gaps 1;

Db 1677 AGARAGYTTKTAARKMYWAMTTMMKMMWTTTWKMMKMYTWTSSRTTTT 1736
      |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Oy 12 AGAAATCTAGTTCACCTCTAATTTTAGTAGTAATAATGAGAGTAATGTTT 71

Db 1737 SAMMMWTWSTWTVYVWVAIYAMKMMWTRTAKAMAMASVARKMKW-TSAAAAYSARKMKW 1795
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Oy 72 CAAGAATCCACCTGCTCCCTAATGCAATTAATAATATTTTAAATACTTTAT 131

Db 1796 WMAVRAKMTMMMAAMKMKRKAAMTWMMWYM 1827
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Oy 132 TTATACTTTTAGAACAATGATGTTGTTA 163

RESULT 2 AF012089 10772 bp DNA INV 05-AUG-1997
LOCUS Drosophila melanogaster cysteine proteinase-1 (Cp1) gene, complete
DEFINITION

```

```

ACCESSION      AF012089
NID            92305220
KEYWORDS
SOURCE
ORGANISM
      Drosophila melanogaster
      fruit fly.
      Eukaryotes; mitochondrial eukaryotes; Metazoa; Arthropoda;
      Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;
      Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
      1 (bases 4546 to 4553)
      Gray,Y.H., Tanaka,M.M. and Sved,J.A.
      P-element-induced recombination in Drosophila melanogaster: hybrid
      element insertion
      Genetics 144 (4), 1601-1610 (1996)
      97132596
      2 (bases 1 to 10772)
      Gray,Y.H.M., Sved,J.A., Preston,C.R. and Engels,W.R.
      Structure of the cysteine proteinase (Cp1) gene of Drosophila
      melanogaster and associated mutational effects
      Unpublished
      3 (bases 1 to 10772)
      Gray,Y.H.M., Sved,J.A., Preston,C.R. and Engels,W.R.
      Direct Submission
      Submitted (30-JUN-1997) School of Biological Sciences, University
      of Sydney, Biology A12, Sydney University, NSW 2006, Australia
      Location/Qualifiers
      1..10772
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      join(872..1000,2310..2426,6476..6690,6751..7707)
      /gene="Cp1"
      /product="cysteine protease"
      872..7707
      /gene="Cp1"
      872..1000
      /gene="Cp1"
      /number=1
      1001..2309
      /gene="Cp1"
      2310..2426
      /gene="Cp1"
      /number=2
      join(2328..2426,6476..6690,6751..7462)
      /gene="Cp1"
      /codon_start=1
      /product="cysteine proteinase-1"
      /db_xref="PID:g2305221"
      /translation="MRTAVLILPLIALLAVAQAVSFADVWMEWHTFKLEHRKNYODT
      EERFKRIFENKHKIAKHNQRPAREKVSFKLAVNKADLIHHEFRLANGFNITLHK
      QLRADSEKGVTFISPAHYTLPSVDMRKGAVTAVKDGHCSCWAFSSSTALEGO
      HFRKSGVLSLSQNLVDCSTKYGNCGNLMDNARYIKDNGIDITEKSEYEAID
      DSCHPFKGVATDRCGFTDIPQDGEKMAAVATGVSVAIDASHESFPQYSEGVAN
      EPCQDAONLDHGVLVGFTGDESGEDYWLKNSMGTWGDGKFIKMLRNKNCGLIAS
      ASSYPLV"
      2427..6475
      /gene="Cp1"
      4546..4553
      /gene="Cp1"
      /note="insertion site of P(casper)(50C)"
      /citation=[1]
      6476..6690
      /gene="Cp1"
      /number=3
      6691..6750
      /gene="Cp1"
      6751..7707
      /gene="Cp1"
      /number=4
      join(<8110..9300,9370..>9532)
      /product="phenylalanyl tRNA synthetase"
      join(8110..9300,9370..>9532)
      /note="potential orf"
      /codon_start=1

```


BASE COUNT		2929 a	2357 c	2282 g	3046 t	158 others
ORIGIN						
Query Match	15.2%	Score 45:	DB 18:	Length 10772:		
Best Local Similarity	22.3%	Pred. No. 5,50e+08:				
Matches	29:	Conservative	61:	Mismatches 39:	Indels	1; Gaps 1;
Db	1719	YRTWMKMYNTSRITTSAMMMWTNSTWIKYMMAYAMKMMWTRTARAKMASWARK	1778			
Cp	228	TAATCTTCATTCATTCATTGCGCTATTGACACTCCAAATGAAGAAAAATCAGAGAACAAATG	169			
Db	1779	WK-TSAAAYASMRKMMWMAVRAKKKMMMAAMKMMKAAWTRMMYMTTAAABRAG	1837			
Cp	168	ATGTTTAACANACTCTACATGTTCTTAAAGATTATTAAGTTATTAATAAATAATA	109			
Db	1838	CTTCTATTGC	1847			
Cp	108	TTTTTAATGC	99			
RESULT	3					
LOCUS	AAU225963	493 bp	DNA	VRT	24-FEB-1998	
DEFINITION	Anguilla anguilla mitochondrial DNA for tRNA-Thr, tRNA-Pro, and					
ACCESSION	AJ225963					
NID	92911583					
KEYWORDS	D-loop; mitochondrion; tRNA-Pro; tRNA-Thr.					
SOURCE	European eel.					
ORGANISM	Mitochondrion Anguilla anguilla					
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;					
AUTHORS	Neopterygii; Teleostei; Anguilliformes; Anguillidae; Anguilla.					
TITLE	1 (bases 1 to 493)					
JOURNAL	Lintas,C., Hirano,J. and Archer,S.N.					
REFERENCE	Genetic variation of the European eel, Anguilla anguilla					
AUTHORS	Unpublished					
TITLE	2 (bases 1 to 493)					
JOURNAL	Archer,S.N.					
REFERENCE	Direct Submission					
AUTHORS	Submitted (22-FEB-1998) Archer S.N., International Marine Centre,					
TITLE	Localita sa Mardini, 09072 Torregrande, Oristano, ITALY					
JOURNAL	Location/Qualifiers					
FEATURES	1..493					
SOURCE	/organism="Anguilla anguilla"					
	/mitochondrion					
	/isolate="RM2 (TIBER RIVER, ROME, ITALY)"					
	/db_xref="taxon:7936"					
	/dev_stage="post-larva"					
	/tissue_type="liver"					
	/tissue_type="muscle"					
	<1..32					
	/gene="tRNA-Thr"					
	1..32					
	/gene="tRNA-Thr"					
	38..107					
	/gene="tRNA-Pro"					
	38..107					
	/gene="tRNA-Pro"					
	108..>493					
BASE COUNT	216 a	84 c	46 g	147 t		
ORIGIN						
Query Match	10.1%	Score 30:	DB 20:	Length 493:		

	Best Local	Similarity 67.4%	Pred No. 2.84e-01:	
	Matches 58;	Conservative 0;	Mismatches 28;	Indels 0; Gaps 0;
Db	252	AAATATGTAAAAATAAATAATATGTAATATTATTCACATTTTACAATAATTTATATAATA	311	
Cp	188	AAAATACTGAAGAACCAATCATGTGTTAAACCAATCTACATGTTTCTAAAAGTATATAATA	129	
Db	312	AACATTAACAAGAATTAAAAACATTTA	337	
Cp	128	AAGTTATTTAATTAATAATAATATTTTA	103	
RESULT	LOCUS	CDZ79651	457 bp	DNA - UNA
DEFINITION		C.deichmanni mitochondrial 16S rRNA gene.		01-OCT-1996
ACCESSION		279651		
NID		g1617279		
KEYWORDS		16S ribosomal RNA; 16S rRNA gene.		
SOURCE		Celaca deichmanni.		
ORGANISM		Mitochondrion Celuca deichmanni		
		Eukaryota; Animalia; Metazoa; Arthropoda; Crustacea; Malacostraca		
		Eucarida; Decapoda; Plecotenata; Brachyura; Brachyhyncha;		
		Ocyropoidea; Ocyropidae; Celuca.		
REFERENCE		1 (bases 1 to 457)		
AUTHORS		Levinson,J., Sturmawer,C. and Christy,J.		
TITLE		Molecular data and biogeography: resolution of a controversy over evolutionary history of a pan-tropical group of invertebrates		
JOURNAL		J. Exp. Mar. Biol. Ecol. (1996) In press		
AUTHORS		2 (bases 1 to 457)		
TITLE		Sturmawer,C., Christy,J. and Levinson,J.		
REFERENCE		Molecular phylogeny analysis of fiddler crabs: a test of the hypothesis of increasing behavioral complexity in evolution		
AUTHORS		Proc. Natl. Acad. Sci. U.S.A. (1996) In press		
JOURNAL		3 (bases 1 to 457)		
AUTHORS		Sturmawer,C.F.		
TITLE		Direct Submission		
DEFINITION		Submitted (29-AUG-1996) Christian F Sturmawer, Department of Zoology, University of Innsbruck, Technikerstrasse 25, Innsbruck, Tirol, A-6020, Austria		
FEATURES		Location/Qualifiers		
source		1..457		
		/organism="Celuca deichmanni"		
		/mitochondrion		
		/isolate="U95, Panama"		
		/tissue-type="muscle from claw or leg"		
		<1..>457		
rRNA		/gene="16S rRNA"		
		/product="16S ribosomal RNA"		
gene		1..457		
		/gene="16S rRNA"		
BASE COUNT		164 a 49 c 68 g 176 t		
ORIGIN				
		Query Match 9.8%; Score 29; DB 32; Length 457;		
		Best Local Similarity 83.7%; Pred. No. 7.32e-01;		
Matches		36; Conservative 0; Mismatches 7; Indels 0; Caps 0;		
Db	185	TAAAGCTTATAAATATTTTATTTATTTACATTTTTTATAA	227	
OY	95	TAAATGCATTAATAAATATTTTATTTAATAAATCTTATTATAA	137	
RESULT	LOCUS	166494	7218 bp	DNA PAT
DEFINITION		Sequence 14 from patent US 5670367.		23-DEC-1997
ACCESSION		166494		
NID		g2724471		
KEYWORDS		.		
SOURCE		Unknown.		
ORGANISM		Unclassified.		
REFERENCE		1 (bases 1 to 7218)		
AUTHORS		Dorner,F., Schefflingner,F. and Falkner,F.Gunter.		

:

:

SRS	/db.xref="dbSTS:G099335"
	199463.1.199572
	/note="9824, WI-3555, Chr. 16, Homo sapiens"
	/db_xref="dbSTS:G04338"
BASE COUNT	60960 a 51778 c 49172 g 53987 t 124 others
ORIGIN	
Query Match	9.8%; Score 29; DB 26; Length 216021;
Best Local Similarity	18.4%; Pred. No. 7.32e-01;
Matches	18; Conservative 48; Mismatches 31; Indels 1; Gaps 1;
Db	1476 KKKKYYCYCYCYCYCYCYCMGRAMAAAWYKRRCSCAMWYMGRARKTYMMARRRGC 1535
Cp	213 TTTCGCCATATTGCACCTCCAAAATAAATAATCAGACAGACAATAAGCATCTTTAAACAATAC 154
Db	1536 ARSYKKKYAMTAAW-TTWCAAAAAAAMTTYMMMY 1572
Cp	153 TACATGTCTTAAGAATTATMAATAAGTTATTTTAAAT 116
RESULT	8
LOCUS	AJ225966 492 bp DNA VRT 24-FEB-1998
DEFINITION	Anguilla anguilla mitochondrial DNA for trNA-Thr, trNA-Pro, and D-loop, isolate RM5.
ACCESSION	AJ225966
NID	G2911598
KEYWORDS	D-loop; mitochondrion; trNA-Pro; trNA-Thr.
ORGANISM	European eel. Mitochondrion Anguilla anguilla Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae; Anguilla. 1 (bases 1 to 492) Lintas,C., Hirano,J. and Archer,S.N. Genetic variation of the European eel, Anguilla anguilla
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 492)
TITLE	Archer,S.N.
JOURNAL	Direct Submission
JOURNAL	Submitted (23-FEB-1998) Archer S.N., International Marine Centre, Locatiga sa Maddini, 09072 Torregrande, Oristano, ITALY
FEATURES	Location/Qualifiers
Source	1..492 /organism="Anguilla anguilla" /mitochondrion /isolate="RM5 (TIBER RIVER, ROME, ITALY)" /db_xref="taxon:7936" /dev_stage="post-larva" /tissue_type="liver" /tissue_type="muscle" <1..32 /gene="trNA-Thr" 1..32 /gene="trNA-Thr" 38..107 /gene="trNA-Pro" 38..107 /gene="trNA-Pro" 108..>492
D-loop	215 a 86 c 45 g 146 t
BASE COUNT	
ORIGIN	
Query Match	9.5%; Score 28; DB 20; Length 492;
Best Local Similarity	66.3%; Pred. No. 1.86e+00;
Matches	57; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
Db	252 AAAATATGTAAAAATCAATAATAGTAAATATATTCATTTACAAAAATTATATAATA 311
Cp	188 AAATATACGAAGAACAAATGATGTTTAAACAATACATGTTCTTAAAGTTATAATA 129
Db	312 AACATAACAAGTTAAAAACAATT 337
Cp	128 AAGTATTTAAATAAATATTTT 103

RESULT	9	AAJ25955	493 bp	DNA	VRT	24-FEB-1998
LOCUS						
DEFINITION		Anguilla anguilla mitochondrial DNA for tRNA-Thr, tRNA-Pro, and D-loop, isolate M3.				
ACCESSION		AAJ25955				
NID		92911575				
KEYWORDS		D-loop; mitochondrion; tRNA-Pro; tRNA-Thr.				
SOURCE		European eel.				
ORGANISM		Mitochondrion Anguilla anguilla				
REFERENCE		Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae; Anguilla.				
AUTHORS		1 (bases 1 to 493)				
TITLE		Llantas,C., Hirano,J. and Archer,S.N.				
JOURNML		Genetic variation of the European eel, Anguilla anguilla				
REFERENCE		Unpublished				
AUTHORS		2 (bases 1 to 493)				
TITLE		Archer,S.N.				
JOURNML		Direct Submission				
FEATURES		Submitted (23-FEB-1998) Archer S.N., International Marine Centre, Localita sa Mardini, 09072 Torregrande, Oristano, ITALY				
SOURCE		Location/Qualifiers				
		1..493				
		/organism="Anguilla anguilla"				
		/mitochondrion				
		/isolate="M3 (CABRAS LAGOON, ORISTANO, ITALY)"				
		/db_xref="taxon:7936"				
		/dev_stage="post-larva"				
		/tissue_type="liver"				
		/tissue_type="muscle"				
		<1..32				
		/gene="tRNA-Thr"				
		1..32				
		/gene="tRNA-Thr"				
		38..107				
		/gene="tRNA-Pro"				
		38..107				
		/gene="tRNA-Pro"				
		108..>493				
		D-loop				
		215 a				
		87 c				
		46 g				
		145 t				
		BASE COUNT				
		ORIGIN				
		Query Match				
		Best Local Similarity 66.3%; Pred. No. 1.86e+00;				
		Matches 57; Conservative 0; Mismatches 29; Indels 0; Gaps 0;				
		Score 28; DB 20; Length 493;				
		9.5%;				
		252 AAAATATGTAATAATCATATATATGTAATAATATATATCCAACTGCACAAATTTATAATA 311				
		CP 188 AAAATATGTAATAATCATATATATGTAATAATATATATCCAACTGCACAAATTTATAATA 129				
		312 AACATACCAAGTTAATAAATCATTTA 337				
		DB 128 AAGTTATTTAATAATAAATATTTT 103				
		RESULT 10				
		AAJ226002				
		LOCUS				
		DEFINITION				
		Anguilla anguilla mitochondrial DNA for tRNA-Thr, tRNA-Pro, and D-loop, isolate GRS.				
		ACCESSION				
		AAJ226002				
		NID				
		92911634				
		KEYWORDS				
		D-loop; mitochondrion; tRNA-Pro; tRNA-Thr.				
		SOURCE				
		European eel.				
		Mitochondrion Anguilla anguilla				
		Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae; Anguilla.				
		1 (bases 1 to 493)				
		Llantas,C., Hirano,J. and Archer,S.N.				
		Genetic variation of the European eel, Anguilla anguilla				
		Unpublished				
		2 (bases 1 to 493)				
		Archer,S.N.				
		Direct Submission				
		Submitted (23-FEB-1998) Archer S.N., International Marine Centre, Localita sa Mardini, 09072 Torregrande, Oristano, ITALY				
		Location/Qualifiers				
		1..493				
		/organism="Anguilla anguilla"				
		/mitochondrion				
		/isolate="M3 (CABRAS LAGOON, ORISTANO, ITALY)"				
		/db_xref="taxon:7936"				
		/dev_stage="post-larva"				
		/tissue_type="liver"				
		/tissue_type="muscle"				
		<1..32				
		/gene="tRNA-Thr"				
		1..32				
		/gene="tRNA-Thr"				
		38..107				
		/gene="tRNA-Pro"				
	</					

Tue Dec 22 09:17:19 1998

US-08-530-112A-563.rge

Page 10

```
Query Match      9.5%; Score 28; DB 15; Length 10929;
Best Local Similarity 63.2%; Pred. No. 1.86e+00;
Matches 67; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

Qy	161	T T A A C A T C A T T T G T T C T T C A G A T T T T T C A T T T G G A A T C C A A T A	206
Db	2634	G G A T A G A T G A T G C A T C A A C A C A A T T C C C T G G A A T T T T A A A	2679
Qy	101	A T T A A A A A A T T T T T A T T A A A T A C T T T A T T A T T A T T A C C T T T A G A A C A T G A T A T T G T	160
Db	2574	A T T A C A A A T C A T T T T T A A C A A T A A C T T A T T A T T A C T T T T T G A A G C T T C C A T C T G	2633

```
Search completed: Tue Dec 22 03:58:07 1998
Job time : 489 secs.
```


Suita, Osaka, 565
Japan
Phone: 06-877-5111
Fax : 06-875-1922
DDBJ CHROMOSOMAL ASSIGNMENT
:method=PCR WITH HYBRID CELL PANEL'
:primer_f='CCGAGGTTCTTGTCTTA'
:primer_r='TCTGACTTGGTCCAG'
:background='mouse:1,2,5,6,7,8,10,11,12,14,15,16,17,18,19,
20,21,X,Y,Chinese hamster:3,4,9,13,22,Y'
:background='chinese hamster:3,4,9,13,22,Y'
:chromosome='6'
:chromosome='12'
Location/Qualifiers
1. 147
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HL60"
/cell_type="promyelocyte"
/dev_stage="adult"
/sex="female"

BASE COUNT 39 a 39 c 30 g 39 t

ORIGIN
Query Match 100.0%; Score 147; DB 7; Length 147;
Best Local Similarity 100.0%; Pred. No. 4,47e-212;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GATCCCCAGGGTTCTTGTCTTATTTATGAGAAAAACCGGTCTTGTCCAGCGAC 60
|||||
1 GATCCCCAGGGTTCTTGTCTTATTTATGAGAAAAACCGGTCTTGTCCAGCGAC 60
|||||

Db 61 TGTGAGCCCCCAGCTCAGCCAGCCCTGGCCCTGGTACTTGAACGAAGTTACA 120
|||||
61 TGTGAGCCCCCAGCTCAGCCAGCCCTGGCCCTGGTACTTGAACGAAGTTACA 120
|||||

Db 121 GATTATATTAATAATATATGACAAA 147
|||||
121 GATTATATTAATAATATATGACAAA 147
|||||

RESULT 2 A1018794 349 bp mRNA EST 16-JUN-1998
LOCUS OY32909.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1639072
DEFINITION 3', mRNA sequence.
ACCESSION A1018794
NID G3232592
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 349)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
www.bio.lnlnl.gov/dbfp/image/image.html

FEATURES
Seq primer: -40m13 fwd. ET from Amersham.
Location/Qualifiers
1. 349
/organism="Homo sapiens"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5].
TCTTCCAACTCTGAAGTGGAGCGGCCGCCCAATTTTTTTTTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

/db_xref="taxon:9606"
/clone_image="IMAGE:1639072"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"

BASE COUNT 114 a 70 c 74 g 91 t

ORIGIN
Query Match 100.0%; Score 147; DB 15; Length 349;
Best Local Similarity 100.0%; Pred. No. 4,47e-212;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 4 TTTGTACATATTTATTTATATATATCTGACTGCTCCAAAGTCAAGGGGGGCC 63
|||||
CP 147 TTTGTACATATTTATTTATATATATCTGACTGCTCCAAAGTCAAGGGGGGCC 88
|||||

Db 64 AGGCTGCGCTGAGTGGGGCTCAGTCAGTCGCTGCAAAAGTGACCGCTTTTCTCAT 123
|||||
CP 87 AGGCTGCGCTGAGTGGGGCTCAGTCAGTCGCTGCAAAAGTGACCGCTTTTCTCAT 28
|||||

Db 124 AAATPAGCAAGAAACCTGGGGATC 150
|||||
CP 27 AAATPAGCAAGAAACCTGGGGATC 1
|||||

RESULT 3 A1017469 478 bp mRNA EST 27-AUG-1998
LOCUS 0u23c08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:1627118 3', mRNA sequence.
ACCESSION A1017469
NID G3231805
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 478)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through ILNL; contact the
IMAGE Consortium (info@image.lnlnl.gov) for further information.
Insert Length: 566 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 386.
Location/Qualifiers
1. 478
/organism="Homo sapiens"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBH119W, testis NHT, and B-cell
NCI-CCAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of

CP 27 AATAAGACAAGAACCTGGGGATC 1

RESULT 6 AA972448 313 bp mRNA EST 07-JUL-1998
LOCUS Op1b02.s1 Soares_NFL_T_GBC_S1 Homo sapiens CDNA clone
DEFINITION IMAGE:1579371 3', mRNA sequence.
ACCESSION AA972448
NID 93145212
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 313)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 419 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.

FEATURES
source
Location/Qualifiers
1..313
/organism="Homo sapiens"
/note="Organ: pooled; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NMT, and B-cell
NCL1CGAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_id="IMAGE:1579371"
/lab_host="DH10B"
/lab_host="DH10B"

BASE COUNT 104 a 59 c 70 g 80 t

ORIGIN
Query Match 95.2%; Score 140; DB 15; Length 313;
Best Local Similarity 98.6%; Pred. No. 2,066-199;
Matches 142; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 1 GTACATTATTTTATATATATCTGTAACCTCGGTTCCAAAGTACCAAGGGGGGCCAGG 60
|||||
CP 144 GTACATTATTTTATATATATCTGTAACCTCGGTTCCAAAGTACCAAGGGGGGCCAGG 85
|||||
Db 61 GCTGGCTCGAGTGGGGCCTCACAGTGGCTGGACAAGTAGTACCGGTTTCTCATATAA 120
|||||
CP 84 GCTGGCCTGAGTGGGGCCTCACAGTGGCTGGACAAGTAGTACCGGTTTCTCATATAA 25
|||||
Db 121 TAAGACAAGAAACCTGGGGATC 144
|||||
CP 24 TAAGACAAGAAACCTGGGGATC 1

RESULT 7 AA862961 232 bp mRNA EST 11-MAR-1998
LOCUS Og98f02.s1 NCI_CGAP_Kid3 Homo sapiens CDNA clone IMAGE:1456347 3',
DEFINITION mRNA sequence.
ACCESSION AA862961
NID 92955440
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 232)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.llnl.gov/btrp/image/image.html
Seq primer: -40m13 fwd. ET from Amersham.

FEATURES
source
Location/Qualifiers
1..232
/organism="Homo sapiens"
/note="Organ: kidney; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified p773 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_id="IMAGE:1456347"
/clone_id="NCI_CGAP_Kid3"
/lab_host="DH10B"

BASE COUNT 68 a 45 c 57 g 62 t

ORIGIN
Query Match 91.8%; Score 135; DB 13; Length 232;
Best Local Similarity 97.3%; Pred. No. 2,146-199;
Matches 144; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Db 1 TTGTACATTATTTTAAATATATCTGTAACCTCGGTTCCAAAGTACCAAGGGGGGCC 60
|||||
CP 147 TTGTACATTATTTTAAATATATCTGTAACCTCGGTTCCAAAGTACCAAGGGGGGCC 88
|||||
Db 61 AGGCTAGGCTGAGTGGGGCCTCACAGTGGCTGGACAAGTAGTACCGGTTTCTCCA 120
|||||
CP 87 AGGGCTGGCCTGAGT-GGGGGCCTCACAGTGGCTGGACAAGTAGTACCGGTTTCTCCA 29
|||||
Db 121 TAAATAGACAAGAAACCTGGGGATC 148
|||||
CP 28 TAAATAGACAAGAAACCTGGGGATC 1

RESULT 8 AA923630 318 bp mRNA EST 29-APR-1998
LOCUS ng98c09.s1 NCI_CGAP_Co10 Homo sapiens CDNA clone IMAGE:1160368 3',
DEFINITION mRNA sequence.
ACCESSION AA923630
NID 93070939
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 318)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Ian Kirsch, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

FEATURES

Insert Length: 422 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 269.
Location/Qualifiers

1..318
/organism="Homo sapiens"
/note="Organ: COLON; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
RER+ colon tumor, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo (N-Soares4)."
/db_xref="taxon:9606"
/clone="IMAGE:1160368"
/tissue_type="colon tumor RER+"
/lab_host="DH10B"

BASE COUNT 105 a 61 c 70 g 82 t
ORIGIN

Query Match 91.2%; Score 134; DB 14; Length 318;
Best Local Similarity 96.5%; Pred. No. 1,35e-188;

Matches 139; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1 GTCATTATTTATTTATATCTGTAATCTGCTGCTCAAGTACCAAGGGGGGCCAGG 60
|||||
Cp 144 GTACATTATTTATTTATATCTGTAATCTGCTGCTCAAGTACCAAGGGGGGCCAGG 85
|||||
Db 61 GCTGGCTGAGATGGGGCCCTCACAGTGCCTGACAAAGTACCGGTTTCTCCATAAA 120
|||||
Cp 84 GCTGGCTGAGTGGGGCCCTCACAGTGCCTGACAAAGTACCGGTTTCTCCATAAA 25
|||||
Db 121 TAAGACAAAGAAACCCCTGGGGATC 144
|||||
Cp 24 TAAGACAAAGAAACCCCTGGGGATC 1

RESULT 9
LOCUS AA557377 273 bp mRNA EST 09-SEP-1997
DEFINITION n181f07.s1 NCI_CGAP_Br2 Homo sapiens CDNA clone IMAGE:1057093, mRNA
sequence.
ACCESSION AA557377
NID 92327854
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 273)
REFERENCE
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

FEATURES

source

Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 2819 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 240.
Location/Qualifiers

1..273
/organism="Homo sapiens"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from pooled bulk
breast tumor tissue, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. This library is the normalized version of
NCI CGAP Br1.1. Library was constructed by Bento Soares
and M. Fatima Bonaldo.
/db_xref="taxon:9606"
/clone="IMAGE:1057093"
/clone_lib="NCI_CGAP_Br2"
/sex="female, pooled"
/tissue_type="breast"
/lab_host="DH10B"
<1..>273

BASE COUNT 97 a 53 c 55 g 68 t
ORIGIN

Query Match 89.8%; Score 132; DB 9; Length 273;
Best Local Similarity 96.5%; Pred. No. 5.37e-185;

Matches 137; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1 ACATTATTTATTTATATATCTGTAATCTGCTGCTCAAGTACCAAGGGGGGCCAGG 60
|||||
Cp 144 ACATTATTTATTTATATATCTGTAATCTGCTGCTCAAGTACCAAGGGGGGCCAGG 83
|||||
Db 61 TGCTCTGAGATGGGGCCCTCACAGTGCCTGACAAAGTACCGGTTTCTCCATAAA 120
|||||
Cp 82 TGCTCTGAGTGGGGCCCTCACAGTGCCTGACAAAGTACCGGTTTCTCCATAAA 23
|||||
Db 121 AGACAAAGAAACCCCTGGGGATC 142
|||||
Cp 22 AGACAAAGAAACCCCTGGGGATC 1

RESULT 10
LOCUS H41523 390 bp mRNA EST 31-JUL-1995
DEFINITION ym90g09.r1 Homo sapiens CDNA clone 175744 5'.
ACCESSION H41523
NID 9917575
KEYWORDS EST.
SOURCE Homo sapiens clone-175744 library-Soares adult brain N25HB55X
vector-pT73D (Pharmacia) with a modified polylinker host-DH10B
(ampicillin resistant) primer-M13P1 Rsite1-Not I Rsite2-Eco RI
55-year old male. 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCGGCTTTTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I and Eco
RI sites of a modified pT73 vector (Pharmacia). Library went
through one round of normalization to a Cot = 53. Library
constructed by Bento Soares and M. Fatima Bonaldo. The adult brain
RNA was provided by Dr. Donald H. Gillen. Tissue was acquired 17-18
hours after death which occurred in consequence of a ruptured
aortic aneurysm. RNA was prepared from a pool of tissues
representing the following areas of the brain: frontal, parietal,

Homo sapiens
Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata;
Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Deuterostomia; Chordata; Tetrapoda; Amniota; Mammalia; Theria;
Sarcopreyalli; Choonata; Tetrapoda; Amniota; Mammalia; Theria;
Sauria; Archonta; Primates; Catarrhini; Hominiidae; Homo.

University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mscoates@blue.weeg.uiowa.edu

oligo-dt track (not shown) served to identify it as a clone from the normalized adult Spleen library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics
Seq primer: M13 Forward.

FEATURES
source
Location/Qualifiers
1..360

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: Eco RI; The UI-R-A1 library is a subcloned library derived from the UI-R-A0 library. The UI-R-A0 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dt track which allows identification of the library of origin of a clone within the mixture. The subcloned library (UI-R-A1) was constructed as follows: PCR amplified cDNA inserts from a pool of approximately 3,840 UI-R-A0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-A0 library in the form of single-stranded circles. The remaining single-stranded circles (subcloned library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-A1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)."
/db_xref="taxon:10116"
/clone="UI-R-A1-eh-b-12-0-UI"
/clone_lib="UI-R-A1"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"

BASE COUNT 122 a 65 c 59 g 114 t
ORIGIN

Query Match 50.3%; Score 74; DB 14; Length 360;
Best Local Similarity 86.7%; Pred. No. 3.46e-83;
Matches 111; Conservative 0; Mismatches 13; Indels 4; Gaps 4;

Db 149 ATTATTTATTTATATATTCGTGTAACGTGGTTCCAGTACCAAGGAGGCCAGCC 208
|||||
Cp 140 ATTATTTATTTATATA-ATCTGTAACCTCGTTCACAGTACCAAGGAGGCCAGCGCT 82
Db 209 GGGCGTGAAGTGGAGCCACAGTCGCGTGGACAGAAAGACTGTTCTT-TCCATTAAT 267
|||
Cp 81 GG-CCAGAGTGGGG-GCCTCACAGTGGCGTGACAAAGTAGTACCGGTTTCTCCATTAAT 24
Db 268 AAGACAAA 275
|||||
Cp 23 AAGACAAA 16

RESULT 13
LOCUS AA122818 310 bp mRNA EST 16-FEB-1997
DEFINITION m03b11.1 Soares mouse 3NDMS Mus musculus cDNA clone 596349 5',
mRNA sequence.
ACCESSION AA122818
NID 91681775
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotes: mitochondrial eukaryotes; Metazoa: Chordata;
Vertebrata: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.

REFERENCE 1 (bases 1 to 310)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

TITLE Waterston.R.
JOURNAL The WashU-HMI Mouse EST Project
COMMENT Unpublished (1996)

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:361781
Putative full length read
vector to vector length is 311
Seq primer: -28M13 rev2 from Amersham.
Location/Qualifiers
1..310

FEATURES
source

/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAGTGGAGCGCGCCGCTGTTTATTTTATTTTATTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Benito Soares and M.Fatima Bonaldo."
/db_xref="taxon:10090"
/clone="596349"
/clone_lib="Soares mouse 3NDMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
<1..>310

BASE COUNT 86 a 55 c 57 g 112 t
ORIGIN mRNA

Query Match 49.0%; Score 72; DB 23; Length 310;
Best Local Similarity 85.9%; Pred. No. 8.26e-80;
Matches 110; Conservative 0; Mismatches 14; Indels 4; Gaps 4;
Db 35 TTGCTTATTATGAAAGAAAT-GGCTTTCGTCCAGCGCACTGAGGCTCCGAC 93
|||||
Oy 16 TTGCTTATTATGAGAAAACCGGCTGCTTGTCCAGCGCACTGAGGCCCCCA-C 74
Db 94 TCAGGCCGCGCTCGCTTCCTTGACTTGGAAACGACGTGACATATATATATAA 153
|||||
Oy 75 TCAGGCCAGCC-CGCGCCCGCCCTTGACTTGGAAACGACGTGACATATATAA 132
Db 154 ATAGTAAT 161
|||
Oy 133 ATATTAAT 140

RESULT 14
LOCUS AA415181 382 bp mRNA EST 16-OCT-1997
DEFINITION vc98e08.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
791078 5', mRNA sequence.
ACCESSION AA415181
NID 62075271
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotes: Metazoa: Chordata: Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 382)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

TITLE
JOURNAL
COMMENT
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
The WashU-HMNI Mouse EST Project
Unpublished (1996)

CONTACT: Marra M/Mouse EST Project
WashU-HMNI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:475486
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 297.
Location/Qualifiers
1..382
/organism="Mus musculus"
/strain="B6D2 F1/J"
/note="Organ: embryo; Vector: Bluescribe (modified);
Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
mRNA prepared from 13,500 2-cell stage embryos. Primer:
SalI(dT): 5'-CGGTGACCGTCGACCGTTTCTTTTCTTTT-3'. CDNAS
were cloned into the MluI/SalI sites of a modified
pBluescribe vector using commercial linkers (NEB).
Average insert size: 1.2 kb."
/db_xref="taxon:10090"
/clone="791078"
/clone_lib="Knowles Solter mouse 2 cell"
/tissue_type="embryo"
/dev_stage="2-cell"
/lab_host="DH10B"

BASE COUNT 111 a 62 c 67 g 142 t
ORIGIN

Query Match 49.0%; Score 72; DB 10; Length 382;
Best Local Similarity 85.9%; Pred. NO. 8.26e-80;
Matches 110; Conservative 0; Mismatches 14; Indels 4; Gaps 4;

Db 105 TTGTCTTATTATGGAAGAAT-GGTCTTCTGTCACGCGCAGCTGTAGGCTCCGAAC 163
|||||
QY 16 TTGTCTTATTATGGAAGAACCGGTCTTCTGTCACGCGCAGCTGTAGGCTCCGAAC 74
|||||
Db 164 TCAGGCCCGGCTGCGCTTCCCTTGTTGTTGTAACGAGCTTACAGATATATATTA 223
|||||
QY 75 TCAGGCCAGCC-CTGGCCCCCCTTGTTGTTGTAACGAGCTTACAGAT-TATATTA 132
|||||
Db 224 ATAGTAAT 231
|||||
QY 133 ATATATAT 140

RESULT 15
LOCUS AA119227 432 bp mRNA EST 17-FEB-1997
DEFINITION mp7b12.r1 Soares 2NBMt Mus musculus cDNA clone 575231 5', mRNA
sequence.
ACCESSION AA119227
NID 91676717
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS
1 (bases 1 to 432)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE
JOURNAL
COMMENT
The WashU-HMNI Mouse EST Project
Unpublished (1996)

CONTACT: Marra M/Mouse EST Project
WashU-HMNI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:349879
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 401.
Location/Qualifiers
1..432
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCGAAGTGGAGGAGCGCGGCTTTTCTTTTCTTTTCTTTT
3)]; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Fatima Bonaldo."
/db_xref="taxon:10090"
/clone="575231"
/clone_lib="Soares 2NBMt"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
<1..>432

BASE COUNT 112 a 79 c 81 g 159 t 1 others
ORIGIN

Query Match 49.0%; Score 72; DB 23; Length 432;
Best Local Similarity 85.9%; Pred. NO. 8.26e-80;
Matches 110; Conservative 0; Mismatches 14; Indels 4; Gaps 4;
Db 245 TTGTCTTATTATGGAAGAAT-GGTCTTCTGTCACGCGCAGCTGTAGGCTCCGAAC 303
|||||
QY 16 TTGTCTTATTATGGAAGAACCGGTCTTCTGTCACGCGCAGCTGTAGGCTCCGAAC 74
|||||
Db 304 TCAGGCCCGGCTGCGCTTCCCTTGTTGTTGTAACGAGCTTACAGATATATATTA 363
|||||
QY 75 TCAGGCCAGCC-CTGGCCCCCCTTGTTGTTGTAACGAGCTTACAGAT-TATATTA 132
|||||
Db 364 ATAGTAAT 371
|||||
QY 133 ATATATAT 140

Search completed: Tue Dec 22 04:16:55 1998
Job time : 262 secs.

 W O R L D
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MSPrch_n n.a. - n.a. database search, using Smith-Waterman algorithm
 Run on: Tue Dec 22 04:17:14 1998; MasPar time 39.01 Seconds
 512.738 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-530-112A-709
 Description: (1-147) from US08530112A.seq
 Perfect Score: 147
 N.A. Sequence: 1 GATCCCCAGGGTTCTTGT.....TTAAATAATATGACAAA 147
 Comp: CTAGGGGTCCTCCAAAGAACAA.....AATTTATATATACATGTTT

Scoring table: TABLE default
 Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: n-geneseg32
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39 40:part40

Statistics: Mean 7.328; Variance 4.796; scale 1.528

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description	Pred. No.
1	147	100.0	147.18	T19709 Human gene signature	1.87e-75
2	40	27.2	91.9	O51746 Oligonucleotide probe	3.66e-10
3	37	25.2	204.1	N81164 Base substituted E.co	1.64e-08
4	36	24.5	204.1	N81164 Base substituted E.co	5.76e-08
5	34	23.1	91.9	O51746 Oligonucleotide probe	6.92e-07
6	26	17.7	250.32	T76438 Substance P antisense	9.81e-03
7	25	17.0	114.12	O70467 Generic DNA sequence	3.07e-02
8	25	17.0	114.12	O70465 Generic DNA sequence	3.07e-02
9	25	17.0	128.32	T76233 Human IL6 antisense o	9.46e-02
10	24	16.3	114.12	O70469 Generic DNA sequence	9.46e-02
11	24	16.3	114.12	O70467 Generic DNA sequence	9.46e-02
12	24	16.3	114.12	O70468 Generic DNA sequence	9.46e-02
13	23	15.6	89.32	T76219 Human IL5 antisense o	2.87e-01

14	23	15.6	114.12	O70470 Generic DNA sequence	2.87e-01
15	23	15.6	114.12	O70468 Generic DNA sequence	2.87e-01
16	23	15.6	114.12	O70473 Generic DNA sequence	2.87e-01
17	23	15.6	114.12	O70465 Generic DNA sequence	2.87e-01
18	22	15.0	114.12	O70466 Generic DNA sequence	8.55e-01
19	22	15.0	114.12	O70472 Generic DNA sequence	8.55e-01
20	22	15.0	114.12	O70466 Generic DNA sequence	8.55e-01
21	22	15.0	3871.2	N71302 HSV-1 gB and surround	8.55e-01
22	21	14.3	74.21	T13613 DC43 TSAR library gen	2.50e+00
23	21	14.3	75.21	T13612 DC43 TSAR library gen	2.50e+00
24	21	14.3	81.21	T13611 DC43 TSAR library gen	2.50e+00
25	21	14.3	82.21	T13610 DC43 TSAR library gen	2.50e+00
26	21	14.3	114.12	O70472 Generic DNA sequence	2.50e+00
27	21	14.3	114.12	O70469 Generic DNA sequence	2.50e+00
28	21	14.3	114.12	O70470 Generic DNA sequence	2.50e+00
29	21	14.3	162.32	T76307 Human RANTES antisens	2.50e+00
30	21	14.3	178.32	T76405 Human endothelin-1 an	2.50e+00
31	21	14.3	250.32	T76438 Substance P antisense	2.50e+00
32	21	14.3	731.5	O31693 RsaI restriction frag	2.50e+00
33	20	13.6	65.21	T13586 TSAR-9 library genera	7.15e+00
34	20	13.6	68.21	T13584 TSAR-9 library genera	7.15e+00
35	20	13.6	74.21	T13613 DC43 TSAR library gen	7.15e+00
36	20	13.6	81.21	T13611 DC43 TSAR library gen	7.15e+00
37	20	13.6	114.12	O70471 Generic DNA sequence	7.15e+00
38	20	13.6	136.32	T76281 Human neutrophil oxid	7.15e+00
39	20	13.6	178.32	T76405 Human endothelin-1 an	7.15e+00
40	20	13.6	190.32	T76452 Chymase antisense o	7.15e+00
41	20	13.6	565.6	O35072 HCV envelope region n	7.15e+00
42	20	13.6	789.2	O12532 cDNA clone encoding h	7.15e+00
43	20	13.6	1856.15	O74089 Human SDF-1-alpha cDN	7.15e+00
44	19	12.9	89.32	T76219 Human IL5 antisense o	2.00e+01
45	19	12.9	2064.3	O14151 Human alpha 2 beta ad	2.00e+01

ALIGNMENTS

RESULT 1
 ID T19709 standard; cDNA to mRNA; 147 BP.
 AC T19709;
 DT 05-JUL-1996 (first entry)
 DE Human gene signature HUMG500783.
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 human; cloning; mapping; non-biased library; diagnosis; detection;
 cell typing; abnormal cell function; ss.
 OS Homo sapiens.
 PN WO9514772-A1.
 PD 01-JUN-1995.
 PE 11-NOV-1994; J01916.
 PR 12-NOV-1993; JP-355504.
 PA (MATSU) MATSUBARA K.
 PI (OKUBO) OKUBO K.
 PI Matsubara K, Okubo K;
 DR WPI: 95-206931/27.
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 for diagnosis of abnormal cell function, by preparing cDNA that
 reflects relative abundance of corresp. mRNA in specific human
 tissues
 PT
 PS Claim 1: Page 454: 2245bp; Japanese.
 CC A single-stranded cDNA (or its complementary strand or the corresp.
 double-stranded DNA) which comprises one of the 7837 "GS" sequences
 given in T19001-T26837 and which is able to hybridise to part of
 human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 sequences were obtained from 3'-directed cDNA libraries prepared
 from various human tissues; synthesis of cDNA was initiated from the
 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 untranslated sequence is unique to a particular mRNA species, almost
 all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 is constructed so as to reflect accurately the relative abundance of
 different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 determined (esp. using primers and probes derived from the GS
 sequences) as a means of diagnosing abnormal cell function or for
 recognising different cell types.

```

SQ Sequence 147 BP; 39 A; 39 C; 30 G; 39 T;
Query Match 100.0%; Score 147; DB 18; Length 147;
Best Local Similarity 100.0%; Pred. No. 1,87e-75;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 gatcccccgggtttcttgcattatattgagaaaacccggtcaactttgccagcgcac 60
      |||||||
Oy 1 GATCCCCGAGGGTTCTTGTCTATTATGTAGGAGAAAAACCGTCACCTTGTCAGCCGCAC 60
      |||||||

Db 61 ttgtgagcccccaactcagcgcaacccttgcccccccttgytacttggaaacgaagttaa 120
      |||||||
Oy 61 TGTGAGCCCCCACCACTCAGGCCAGCCCTTGCCCCCTTGTACTTGGTAACCAAGATTACA 120
      |||||||

Db 121 gattataataataataatgtaaca 147
      |||||||
Oy 121 GATTATTATAAATAATATATGTAACAAA 147
      |||||||

RESULT 2
ID Q51746 standard; cDNA; 91 BP.
AC Q51746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651
PA (BECT ) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
PI WPI: 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC CC reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
CO Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 27.2%; Score 40; DB 9; Length 91;
Best Local Similarity 0.0%; Pred. No. 3.66e-10;
Matches 0; Conservative 44; Mismatches 4; Indels 0; Gaps 0;

Db 13 vhsyyvvvhvshhsvhvhhvsvvvvvhvvhvhyhyvysv 60
      ::::::::::::::::::::
Oy 64 GAGGCCCCACTCAGGCCAGCCCTGAGCCCCCTTGTACTTGAAC 111
      ::::::::::::::::::::

RESULT 3
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
DE E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT /tag= a
FT primer_bind 187..204
FT /function=multiple cloning site
FT /*tag= b
EP-285123-A.
PD 05-MAY-1988.
PE 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
LEhtoavaara P, Knowles J, Koivuola A, Bamford J, Reinikainen T;
```

```

DR      WPI: 88-279927/40.
PR      Introducing random point mutations into nucleic acids -
PT      by prep'n of single stranded template, annealing a primer, elongation,
PS      misincorporation, completion of molecules and screening.
PS      Disclosure: P: English.
CC      Random point mutations were introduced into the alpha fragment of
CC      E.coli beta-galactosidase. The wild type sequence was obtained as a
CC      single stranded template and an oligonucleotide was hybridised to
CC      it to generate a pop'n of DNA molecules which terminate at all
CC      possible nucleotide positions within a specified region. The
CC      variable 3' ends generated in this way are used as primers for
CC      reverse transcriptase. Nucleotides are misincorporated by the
CC      transcriptase and the molecules are completed to forms that can be
CC      amplified and then expressed in a suitable host-vector system.
CC      The sequence covers all 176 diff't base substitutions, most of which
CC      occurred singularly in any given mutant.
CC      See also P80575.
SQ      Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match      25.2%; Score 37; DB 1; Length 204;
Best Local Similarity 8.1%; Pred. No. 1,64e-08;
Matches 8; Conservative 53; Mismatches 38; Indels 0; Gaps 0.

Db      92 hhyrmduvrydyrnsdaawccyrtsvkydcynachdhdyvbybbvynvnhanc 151
      47 TTTGCGCAGCCACTGTGAGGCCGCCCTCAGTCGACCCCTGGCCCCCTTGACTTG 106
      152 nccbnhvhvnhbhnhrwayrvharidvncch 190
      107 GACCGAGATTACGATTATATTAAATAATATGTCACA 145
      108 nccbnhvhvnhbhnhrwayrvharidvncch 190

RESULT 4
ID      N81164 standard; DNA; 204 BP.
AC      N81164.
DT      08-NOV-1990 (first entry)
DE      Base substituted E.coli beta-galactosidase alpha-fragment.
KW      E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS      Escherichia coli.
FH      Key
FT      Location/Qualifiers
      misc_feature 19..69
      tag-a
      /function=multiple cloning site
      primer_bind 187..204
      /tag-b

EP-285123-A.
PD      05-MAY-1988.
PF      30-MAR-1988; 105163.
PR      03-APR-1987; US-034819.
PA      (SUSO) SUDOMEN SOKERI OY.
PI      Lehtoavaara P, Knowles J, Kolvula A, Bamford J, Reinikainen T;
      WPI: 88-279927/40.
PT      Introducing random point mutations into nucleic acids -
PT      by prep'n of single stranded template, annealing a primer, elongation,
PS      misincorporation, completion of molecules and screening.
PS      Disclosure: P: English.
CC      Random point mutations were introduced into the alpha fragment of
CC      E.coli beta-galactosidase. The wild type sequence was obtained as a
CC      single stranded template and an oligonucleotide was hybridised to
CC      it to generate a pop'n of DNA molecules which terminate at all
CC      possible nucleotide positions within a specified region. The
CC      variable 3' ends generated in this way are used as primers for
CC      reverse transcriptase. Nucleotides are misincorporated by the
CC      transcriptase and the molecules are completed to forms that can be
CC      amplified and then expressed in a suitable host-vector system.
CC      The sequence covers all 176 diff't base substitutions, most of which
CC      occurred singularly in any given mutant.
CC      See also P80575.
SQ      Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match      24.5%; Score 36; DB 1; Length 204;
Best Local Similarity 6.2%; Pred. No. 5.76e-08;
Matches 6; Conservative 53; Mismatches 37; Indels 0; Gaps 0.

```

Db	86	ymrththyrmdbvhydyndrdaaayccyrsvkydcocynachddhyvbbvrynn	145
Cp	107	CCAACTTACCAGAGGGGGGCCACAGGCTGCCTGACTGGGGGCTTACAGATGCGCTGGACAA	48
Db	146	hnhnnccecbnhvchvbnhnrwayvhdarr	181
Cp	47	AGTGACCGTTTCTTCATTAATAAGACAAGAA	12

RESULT	5
ID	OS1746 standard; cDNA; 91 BP.
AC	051746:
DT	31-MAY-1994 (first entry)
DE	Oligonucleotide probe MK14-A
KW	Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW	ss.
OS	Synthetic.
PN	EP-571911-A.
PD	01-DEC-1993.
PE	24-MAY-1993: 108325.
PR	26-MAY-1992: US-888651.
PA	(BECT) BECTON DICKINSON CO.
PI	Shank DD, Spears PA;
DR	WPI: 93-378844/48.
PT	New oligo:nucleotide probes specific for Mycobacteria - used for
PT	detection and amplification of Mycobacteria nucleic acid in
PS	samples
PS	Claim 3; Page 14; 23pp. English.
CC	Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC	(051735). It hybridized to all spp. of mycobacteria tested, but
CC	cross reacted to a few non-mycobacterial spp. The probe may
CC	be useful as an initial screen for mycobacterial infection.
CC	See also 051735-45 and 051747-59
SO	Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Db	7	ggcgssvshyvvhvshhshvvhvhhvsvvvhvhhvvhvhhvhyhv	56
Cp	72	GGGGGCGCTCACAGTCGCTGGACAAAGTACC GGCTTTTCCATTAATA	23

RESULT	6
ID	T76438 standard; DNA; 250 BP.
AC	T76438:
DT	16-SEP-1997 (first entry)
DE	Substance P antisense oligonucleotide.
KW	Asthma: airway epithelium; adenose free; cystic fibrosis;
KW	chronic obstructive pulmonary disease; bronchitis; ss.
OS	Synthetic.
PN	WO9640162-A1.
PD	19-DEC-1996.
PE	06-JUN-1996: U09306.
PR	07-JUN-1995: US-474497.
PA	(UYEC-) UNIV EAST CAROLINA.
PI	Meitzger WJ, Nyce JW;
DR	WPI: 97-051871/05.
PT	Treatment of airway diseases such as asthma - by topically applying
PT	adenosine-free antisense oligo:nucleotide to airway epithelium of
PT	subject
PS	Example 5; Page 39; 71pp. English.
CC	A method for treating airway disease in a subject has been produced,
CC	which involves the topical administration of an essentially adenosine
CC	free antisense oligonucleotide (ON) to the airway epithelium of the
CC	subject. The present sequence is an antisense oligonucleotide specific
CC	for the substance P, targeted at the initiation codon. The method can
CC	be used to treat airway diseases such as cystic fibrosis, asthma,
CC	chronic obstructive pulmonary disease, bronchitis and other airway
CC	diseases characterised by an inflammatory response. By eliminating
CC	adenosine from the antisense ON, its liberation upon antisense

	CC	degradation is prevented, thereby preventing adenosine-induced
	CC	bronchoconstriction in patients with hyper-reactive airways.
	SQ	Sequence 250 BP; 1 A; 64 C; 70 G; 65 T;
Dd		Query Match 17.7%; Score 26; DB 32; Length 250; Best Local Similarity 39.7%; Pred. No. 9,81e-03; Matches 22; Conservative 19; Mismatches 16; Indels 0; Gaps 0;
Oy	133	tgcbbcbgctgbytgatgbdbbbbdbtcgcbggcgcgbgbltttcby 130 :: : : : : : : : : : : : 49 TGTCAGGCGACTGTGAGCCGCCCTCAGCGACGCCCTGGCCCCCTTGACTTG 105
RESULT	7	
ID	070467	standard; DNA: 114 BP.
AC	Q70467;	
DR	05-APR-1995	(first entry)
DE	Generic DNA sequence to generate a random TSAR peptide library.	
KW	TSAR: totally synthetic affinity reagent; synthetic; binding domain: effector domain; concatenated heterofunctional protein; linker; direct; rapid; detection; screening; treatment; generic; ss. Synthetic.	
OS	Key	
FH	Location/Qualifiers	
FT	misc_feature	5..60
FT	/tag=	a
FT	/note=	"this sequence represents 'Z'; Z can be a sequence of 6, 9 or 12 nucleotides (see comments)"
FT		
PN	WO9418318-A.	
PD	18-AUG-1994.	
PR	01-FEB-1994;	U00977.
PR	01-FEB-1993;	US-013416.
PR	30-DEC-1993;	US-176500.
PR	31-JAN-1994;	US-189341.
PA	(UYNC-) UNIV NORTH CAROLINA.	
PI	Fowlkes DM, Key BK;	
DR	WPI: 94-279739/34.	
P	P-PsDB; R65153.	
PT	Identifying proteins or peptide(s) which bind a ligand - by screening a recombinant vector library expressing fusion proteins comprising a binding domain and an effector domain	
PS	Disclosure; Page 35; 255pp; English.	
CC	Q70467 is a generic DNA sequence used to generate random TSAR (Totally Synthetic Affinity Reagents) peptides.This generic formula can also be represented as follows: X(NNB)16(TGC)(NNB)12(NNB)16(TGC)(NNB)1Y. X and Y are flanking restriction sites (X is not the same as Y) that are not specified further. Other generic sequences are shown in Q70466-68. Other specific peptides generated by these generic sequences are shown in CC R65151-54. TSARS are concatenated heterofunctional proteins or peptides, comprising at least two functional regions - a binding domain with affinity for a ligand and a second effector peptide portion that is chemically or biologically active.They may further comprise a linker peptide between the 2 domains.The oligonucleotides are also designed so that the expressed peptide contains 2 or 4 cysteine residues positioned in, or flanking, the unpredicted or variant residues. These residues confere some degree of conformational rigidity to the peptides. The TSARS or comppsns comprising a TSAR binding domain can be used in vivo to deliver a chemically or biologically active moiety, eg. metal ion, on radioisotope, peptide, toxin or enzyme, to the specific target or on the cell. They can also replace the function of macromolecules, eg. monoclonal or polyclonal antibodies and therefore circumvent the need for complex methods of hybridoma formation or in vivo antibody production. The TSARS are easily characterised and have designed actively allowing direct and rapid detection in a screening process.	
SQ	Sequence 114 BP;	0 A; 2 C; 2 G; 2 T;
Query Match	17.0%;	Score 25; DB 12; Length 114;
Best Local Similarity	1.0%;	Pred. No. 3.07e-02;
Matches	1; Conservative	28; Mismatches 68; Indels 0; Gaps 0;
Dd	9	bmbnbmbnbnbnbnbnbnbnbnbnbnbnbnbnbnbtgcmbnnnnnnnnbnbn 68
Oy	3	TCCCAGGGTTTCTTGTATTATTAAGAAAAAACCGGTCACTTGTCCAGGCGCAT 62

[illegible][illegible]

CC	comprising at least two functional regions - a binding domain with
CC	affinity for a ligand and a second effector peptide portion that is
CC	chemically or biologically active.They may further comprise a linker
CC	peptide between the 2 domains.The oligonucleotides are also designed so
CC	that the expressed peptide contains 2 or 4 cysteine residues positioned
CC	in, or flanking, the unpredicted or variant residues. These residues
CC	confer some degree of conformational rigidity to the peptides. The TSARs
CC	or compsns. comprising a TSAR binding domain can be used in vivo to
CC	deliver a chemically or biologically active moiety, eg. metal ion,
CC	radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC	cell. They can also replace the function of macromolecules, eg.
CC	monoclonal or polyclonal antibodies and therefore circumvent the need for
CC	complex methods of hybridoma formation or in vivo antibody production.
CC	The TSARs are easily characterised and have designed actively allowing
CC	direct and rapid detection in a screening process.
SQ	Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;
Db	Query Match 16.3%; Score 24; DB 12; Length 114; Best Local Similarity 2.3%; Pzed. No. 9,46e-02; Matches 2; Conservative 25; Mismatches 61; Indels 0; Gaps 0;
Cp	18 bnbhbnnbn 77 TTATTATTTTAAATPACTGTGTAACCTCGTCACAAGTACCAGGGGGGCCAGGCCTGG 80
Db	78 bnbhbnnbn 105
Cp	79 CCTGAGTGCGGGCCTCACAGTGCCTGG 52
RESULT 12	
ID	070468 standard; DNA; 114 BP.
AC	070468:
DT	05-APR-1995 (first entry)
DE	Generic DNA sequence to generate a random TSAR petide library.
KW	TSAR; totally synthetic affinity reagent; sythetic; binding domain;
KW	effector domain; concatenated heterofunctional protein; linker;
KW	direct; rapid; detection; screening; treatment; generic; ss.
OS	Synthetic.
FH	Key Location/Qualifiers
FT	misc_feature 55..60
FT	/tag a
FT	/note= "this sequence represents '2'; 2 can be a
FT	sequence of 6, 9 or 12 nucleotides (see
FT	comments)"
FN	WO9418318-A.
PD	18-AUG-1994.
PR	01-FEB-1994; UO00977.
PR	01-FEB-1993; US-013416.
PR	30-DEC-1993; US-176500.
PR	31-JAN-1994; US-189331.
PA	(UYNC-) UNITV NORTH CAROLINA.
P1	Fowlkes DM, Kay BK;
DR	WPJ; 94-279739/34.
PT	P-PSDB; R65154.
PT	Identifying proteins or peptide(s) which bind a ligand - by
PT	screening a recombinant vector library expressing fusion proteins
PS	disclosure; Page 35; 25pp; English.
CC	Q70468 is a generic DNA sequence used to generate random TSAR (totally
CC	Synthetic Affinity Reagents) peptides.This generic formula can also be
CC	represented as follows: X(NNB)11(TGC)(NNB)62(NNB)7(TGC)(NNB)10Y. X
CC	and Y are flanking restriction sites (X is not the same as Y) that are
CC	not specified further. Other generic sequences are shown in Q70466-68.
CC	Other specific peptides generated by these generic sequences are shown in
CC	R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC	comprising at least two functional regions - a binding domain with
CC	affinity for a ligand and a second effector peptide portion that is
CC	chemically or biologically active.They may further comprise a linker
CC	peptide between the 2 domains.The oligonucleotides are also designed so
CC	that the expressed peptide contains 2 or 4 cysteine residues positioned
CC	in, or flanking, the unpredicted or variant residues. These residues
CC	confer some degree of conformational rigidity to the peptides. The TSARs

PA (UNNC-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Kay BK;
 DR WPI: 94-279739/34.
 P-PSDB: R65154.
 PR Identifying proteins or peptide(s) which bind a ligand - by
 screening a recombinant vector library expressing fusion proteins
 comprising a binding domain and an effector domain
 PS Disclosure: Page 35; 255pp; English.
 CC Q70468 is a generic DNA sequence used to generate random TSAR (Totally
 Synthetic Affinity Reagents) peptides. This generic formula can also be
 represented as follows: X(NNB)11(TGC)(NNB)62(NNB)7(TGC)(NNB)10Y. X
 and Y are flanking restriction sites (X is not the same as Y) that are
 not specified further. Other generic sequences are shown in Q70466-68.
 CC Other specific peptides generated by these generic sequences are shown in
 CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
 CC comprising at least two functional regions - a binding domain with
 CC affinity for a ligand and a second effector peptide portion that is
 CC chemically or biologically active. They may further comprise a linker
 CC peptide between the 2 domains. The oligonucleotides are also designed so
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned
 CC in, or flanking, the unpredicted or variant residues. These residues
 CC confer some degree of conformational rigidity to the peptides. The TSARs
 CC or compps. comprising a TSAR binding domain can be used in vivo to
 CC deliver a chemically or biologically active moiety, eg. metal ion, the
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
 CC cell. They can also replace the function of macromolecules, eg.
 CC monoclonal or polyclonal antibodies and therefore circumvent the need
 CC for complex methods of hybridoma formation or in vivo antibody
 CC production. The TSARs are easily characterised and have designed actively
 CC allowing direct and rapid detection in a screening process.
 SQ Sequence 114 BP. 0 A; 2 C; 2 G; 2 T;

Query Match 15.6%; Score 23; DB 12; Length 114;
 Best Local Similarity 4.4%; Pred. No. 2.87e-01;
 Matches 4; Conservative 24; Mismatches 63; Indels 0; Gaps 0;

DB 24 bnbnnbnnbgtcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbgtg 83
 OY 3 TCCCCAGGGGTTCTTGTATTATGAGAAAAACCGGTCACTTGTCCAGCGCACTG 62
 DB 84 cnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
 OY 63 TGAGGCCCCCCTCACTCAGGCCAGCCCTGCCCC 93

Search completed: Tue Dec 22 04:17:56 1998
 Job time : 42 secs.


```

/clone="119j3"
/chromosome="5"
/note="LBNI H1/5"
Repeat_region
/rpt_family="Alu"
893. .1030
Repeat_region
/rpt_family="Alu"
2295. .2436
Repeat_region
/rpt_family="Alu"
2818. .2859
/note="(G1)21"
/rpt_type=tandem
/rpt_unit=GT
join(3246. .3410,3721. .3828)
/misc_feature
/note="658 & 698 protein identity GenPept:U22377"
Repeat_region
/rpt_family="Alu"
3431. .3724
Repeat_region
/rpt_family="Alu"
3707. .3728
/note="(A)22"
/rpt_type=tandem
/rpt_unit=A
4366. .4661
Repeat_region
/rpt_family="Alu"
5327. .5602
Repeat_region
/rpt_family="Alu"
6586. .6956
Repeat_region
/rpt_family="L1"
6647. .6684
/note="(CA)19"
/rpt_type=tandem
/rpt_unit=CA
7113. .7373
Repeat_region
/rpt_family="Alu"
complement(7830. .8185)
/standard_name="possible repeat"
Repeat_region
8258. .8503
Repeat_region
/rpt_family="Alu"
9070. .9387
Repeat_region
/rpt_family="Alu"
complement(9740. .9845)
Repeat_region
/rpt_family="MER42"
complement(10440. .11015)
Repeat_region
/rpt_family="Alu"
11950. .12250
Repeat_region
/rpt_family="Alu"
12057. .12085
/note="(A)29"
/rpt_type=tandem
/rpt_unit=A
12365. .12645
Repeat_region
/rpt_family="Alu"
13727. .13750
/note="(AC)12"
/rpt_type=tandem
/rpt_unit=AC
13783. .14024
Repeat_region
/rpt_family="L1"
14175. .14470
Repeat_region
/rpt_family="Alu"
complement(14906. .15259)
/standard_name="possible repeat"
Repeat_region
15300. .15613
Repeat_region
/rpt_family="Alu"
16671. .16690
/note="(T)20"
/rpt_type=tandem
/rpt_unit=T
complement(16675. .16977)
Repeat_region
/rpt_family="Alu"
complement(16993. .17085)
Repeat_region
/rpt_family="MER42"
complement(17678. .18276)
Repeat_region
/rpt_family="Alu"
19305. .19583

```

```

Repeat_region
/rpt_family="Alu"
19914. .19945
/note="(GTT)8"
/rpt_type=tandem
/rpt_unit=GTT
complement(19943. .20222)
Repeat_region
/rpt_family="Alu"
21202. .21496
Repeat_region
/rpt_family="Alu"
21736. .22035
Repeat_region
/rpt_family="Alu"
22017. .22038
/note="(A)22"
/rpt_type=tandem
/rpt_unit=A
22321. .22457
Repeat_region
/rpt_family="Alu"
complement(22638. .22981)
/standard_name="possible repeat"
Repeat_region
23022. .23326
Repeat_region
/rpt_family="Alu"
23473. .23761
Repeat_region
/rpt_family="Alu"
23744. .23767
/note="(A)24"
/rpt_type=tandem
/rpt_unit=A
complement(24639. .24694)
Repeat_region
/rpt_family="MER42"
complement(25349. .25713)
Repeat_region
/rpt_family="Alu"
complement(25727. .26471)
Repeat_region
/rpt_family="Alu"
27191. .27477
Repeat_region
/rpt_family="Alu"
27774. .28057
Repeat_region
/rpt_family="Alu"
28040. .28066
/note="(A)27"
/rpt_type=tandem
/rpt_unit=A
28712. .28930
Repeat_region
/rpt_family="MER20"
complement(28769. .28838)
Repeat_region
/note="GRAIL 2 excellent exon, frame 0"
28387. .29214
Repeat_region
/rpt_family="Alu"
join(29485. .29559,29598. .29779)
/note="958 identity dbSTS:G14522 (SHGC-11312)"
29495. .29976
/standard_name="A1027942"
/note="100% identity EST ov84a10.x1"
complement(30401. .30536)
Repeat_region
/note="GRAIL 2 excellent exon, frame 2"
complement(30682. .30733)
Repeat_region
/rpt_family="MIR"
complement(31573. .31724)
/misc_feature
/note="GRAIL 2 excellent exon, frame 1"
complement(32159. .32232)
Repeat_region
/note="GRAIL 2 excellent exon, frame 2"
32368. .32488
Repeat_region
/rpt_family="MLT1"
32617. .32908
Repeat_region
/rpt_family="Alu"
32977. .33088
Repeat_region
/rpt_family="MLT1"
complement(33670. .33785)
Repeat_region
/rpt_family="Alu"
complement(34021. .34144)
/misc_feature
/note="GRAIL 2 excellent exon, frame 2"
complement(35238. .35331)
Repeat_region
/note="GRAIL 2 excellent exon, frame 1"
36392. .36663

```

```

misc_feature      /rpt_family="Alu"
                  36901..37222
                  /note="100% identity EST ou55c09.x1"
                  /db_xref="dbEST:Al025011"
STS
                  36901..37164
                  /standard_name="STSG-9983"
                  /db_xref="dbSTS:G26554"
                  /complement(37595..37654)
misc_feature      /note="GRAIL 2 excellent exon, frame 0"
                  complement(join(38063..38218,38462..38578,38741..38995,
                  39071..39205,39532..39630,39935..40048,40300..40410,
                  40503..40661,41868..41972,42103..42225,42492..42569,
                  44379..44507))
                  /standard_name="histidyl-tRNA synthetase"
                  /note="78% 100% protein identity GenPept:U18937"
misc_feature      /complement(38069..38215)
                  /note="GRAIL 2 excellent exon, frame 0"
misc_feature      complement(38462..38578)

...
Note: remainder of annotations omitted.

Query Match      21.1%; Score 31; DB 26; Length 74371;
Best Local Similarity 18.7%; Pred. No. 1.27e-04;
Matches 20; Conservative 54; Mismatches 31; Indels 2; Gaps 2;

Db 16047 SCMSKMMWTTTTTTTTRTWTTRTKRKRKAGASASRSKSKRCMSYKSKCKKMTCTY-Y 16105
      4 CCCCAGAGGTTTCTTTCTTATATGAGAAACCGGCTCTTCTCCAGCGCACTGT 63
      16106 GMYSMCSYMKYSRYCCGCCYSSYSCSCMMMSMRKKSMMKRM 16152
      64 GAGGCCCCACATCAGGC-CAGCCCTGGCCCCCCTTGTAATTGGA 109

RESULT 2
LOCUS      128278      215 bp      DNA
DEFINITION      Sequence 5 from patent US 5569830.
ACCESSION      I28278
NID          91819054
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 215)
AUTHORS      Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
TITLE      Plant inhibitors of fungal polygalacturonases and their use to
            control fungal disease
JOURNAL      Patent: US 5569830-A 5 29-OCT-1996;
FEATURES
            Source
            1..215
            /organism="unknown"
BASE COUNT      15 a 8 c 25 g 26 t 141 others
ORIGIN
Query Match      19.0%; Score 28; DB 21; Length 215;
Best Local Similarity 17.1%; Pred. No. 6.86e-03;
Matches 18; Conservative 44; Mismatches 42; Indels 1; Gaps 1;

Db 77 KTHYHTNVGSDSKTVDSYSSNGTDCNRSGADYGSSTKMTSRNRTGKT 136
      42 GTCACCTTGTCCAGCCACTGTCAGCCCCCACTGAGCCAGCCCTTGGT 101
      137 ANNAVDSRNMGASVG-SPKNTKKAHKNASADGKVGSKNNDRNR 180
      102 ACTTGGAACCGAAGTACAGATATATTAATAATATATGACAA 146

RESULT 3
LOCUS      HUAC004787 216021 bp      DNA
DEFINITION      Homo sapiens Chromosome 16 BAC clone C19875K-A-952F10, complete
            sequence.
ACCESSION      AC004787
NID          93337381

```

```

KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS      Primates; Catarrhini; Homiidae; Homo.
            1 (bases 1 to 216021)
            Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venner,J.C.
            Homo sapiens Chromosome 16 BAC clone C19875K-A-952F10
REFERENCE      Unpublished
AUTHORS      2 (bases 1 to 216021)
            Adams,M.D. and Loftus,B.J.
TITLE      Direct Submission
JOURNAL      Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA, Email:
            bjoftus@tigr.org
            3 (bases 1 to 216021)
            Adams,M.D. and Loftus,B.J.
TITLE      Direct Submission
JOURNAL      Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
            On Jul 24, 1998 this sequence version replaced g1:3241936.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr., Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org. The orientation of the sequence is from
            sp6 end to 3' end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
            Washington), GenScan (Chris Burge,
            http://genomic.stanford.edu/~chris/GENSCANW.html) searches of the
            complete sequence against a peptide database, and the Human gene
            index database at TIGR (http://www.tigr.org/tdb/hgi.html).
            Genes without peptide homology having spliced EST hits are termed
            'unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES
            Location/Qualifiers
            1..216021
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="16"
            /map="16q21-22"
            /clone="RA-952F10"
            2765..27872
            /note="7766, STS1-CSRL-27g3-uA/CSRL-27g3-uZ, Chr. -, Homo
            sapiens"
            /db_xref="dbSTS:G02280"
            73826..73943
            /note="7608, STS1-CSRL-24g1-uA/CSRL-24g1-uZ, Chr. -, Homo
            sapiens"
            /db_xref="dbSTS:G02122"
            175801..175945
            /note="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"
            /db_xref="dbSTS:G09703"
            175810..175945
            /note="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"
            /db_xref="dbSTS:G09935"
            199463..199572
            /note="9824, WI-3555, Chr. 16, Homo sapiens"
            /db_xref="dbSTS:G04338"
BASE COUNT      60960 a 51778 c 49172 g 53987 t 124 others
ORIGIN
Query Match      19.0%; Score 28; DB 26; Length 216021;
Best Local Similarity 10.6%; Pred. No. 6.86e-03;
Matches 10; Conservative 51; Mismatches 33; Indels 0; Gaps 0;

Db 1448 CSBARRGKKKTKMKSMRRARRSGAKKKKYYYYCYYYCYCMGRAMMAAWYK 1507
      100 CCAAGGGGGGCGCAGCGCTGAGTGCGGCGCTCACAGTGCCTGACAAAGTACC 41
      1508 RRSMAWYYMYGRRKTYMYARRRGARSYK 1541
      40 GGTTCCTTCATTAATAGACAAAGAAACCTG 7

```


STS	/note="1608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo sapiens"	
	/db_xref="dbSTS:G02122"	
	175801..175945	
	/note="16084, ChLC.GCM10B02, Chr. -, Homo sapiens"	
	/db_xref="dbSTS:G09935"	
STS	175810..175945	
	/note="16316, ChLC.GCT1SC04, Chr. -, Homo sapiens"	
	/db_xref="dbSTS:G09572	
	199463..199572	
	/note="9824, WI-3555, Chr. 16, Homo sapiens"	
	/db_xref="dbSTS:G04338"	
BASE COUNT	60960 a	51778 c 49172 g 53987 t 124 others
ORIGIN		
Query Match	17.7%	Score 26; DB 26; Length 216021;
Best Local Similarity	11.4%	Pred. No. 8,91e-02;
Matches	5; Conservative	30; Mismatches 9; Indels 0; Gaps 0;
Db	1475	KKKKRYYCYCYYYYYYCMGRAMMAAAMYYKRRSCNAMYVM 1518
	9	GGGTTCTTGTCTTATATATGAGAAAACGGTCACATTGTC 52
RESULT	8	
LOCUS	AF012089	10772 bp DNA INV 05-AUG-1997
DEFINITION	Drosophila melanogaster cysteine proteinase-1 (Cp1) gene, complete cds, and phenylalanyl tRNA synthetase gene, partial cds.	
ACCESSION	AF012089	
NID	92305220	
KEYWORDS	fruit fly.	
SOURCE	Drosophila melanogaster	
ORGANISM	Eukaryotes; mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 4546 to 4553)	
AUTHORS	Gray,Y.H., Tanaka,M.M. and Svad,J.A.	
TITLE	P-element-Induced recombination in Drosophila melanogaster: hybrid element insertion	
JOURNAL	Genetics 144 (4), 1601-1610 (1996)	
MEDLINE	97132596	
REFERENCE	2 (bases 1 to 10772)	
AUTHORS	Gray,Y.H.M., Svad,J.A., Preston,C.R. and Engels,W.R.	
TITLE	Structure of the cysteine proteinase (Cp1) gene of Drosophila melanogaster and associated mutational effects	
JOURNAL	Unpublished	
REFERENCE	3 (bases 1 to 10772)	
AUTHORS	Gray,Y.H.M., Svad,J.A., Preston,C.R. and Engels,W.R.	
TITLE	Direct Submission	
JOURNAL	Submitted (30-JUN-1997) School of Biological Sciences, University of Sydney, Biology A12, Sydney University, NSW 2006, Australia	
FEATURES	Location/Qualifiers	
source	1..10772	
	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"	
mRNA	join(872..1000,2310..2426,6476..6690,6751..7707)	
	/gene="Cp1"	
	/product="cysteine protease"	
gene	872..7707	
	/gene="Cp1"	
exon	872..1000	
	/gene="Cp1"	
	/number=1	
intron	1001..2309	
	/gene="Cp1"	
	2310..2426	
exon	/gene="Cp1"	
	/number=2	
	join(2328..2426,6476..6690,6751..7462)	
CDS	/gene="Cp1"	
	/codon_start=1	
	/product="cysteine proteinase-1"	

AUTHORS	Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M. Davs,C.A., Kader,K., Mysel,T., Pitluck,S., Pollard,M., Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE	Direct Submission
JOURNAL	Submitted (01-AUG-1998) Human Genome Center, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, MS 74-157, Berkeley, CA 94720, U.S.A.
COMMENT	Sequence submitted by: DOE Joint Genome Institute. Location/Qualifiers
FEATURES	1. 74371
source	/organism="Homo sapiens" /db_xref="taxon:9606" /map="Sq" /clone="119j3" /chromosome="5" /note="BABL H175" 893. 1030 /rpt_family="Alu" 2295. 2438 /rpt_family="Alu" 2818. 2859 /note="(GT)21" /rpt_type=cadem /rpt_unit=GT join(3246. 3410,3721. 3828) /standard_name="RLF" /note="65% & 69% protein identity GenPept:U02377" 3431. 3724 /rpt_family="Alu" 3707. 3728 /note="(A)22" /rpt_type=cadem /rpt_unit=A 4366. 4661 /rpt_family="Alu" 5327. 5602 /rpt_family="Alu" 6586. 6956 /rpt_family="L1" 6647. 6684 /note="(CA)19" /rpt_type=cadem /rpt_unit=CA 7113. 7373 /rpt_family="Alu" complement(7810. 8185) /standard_name="possible repeat" 8258. 8503 /rpt_family="Alu" 9070. 9387 /rpt_family="Alu" complement(9740. 9845) /rpt_family="MER42" complement(10440. 11015) /rpt_family="Alu" 11950. 12250 /rpt_family="Alu" 12057. 12085 /note="(A)29" /rpt_type=cadem /rpt_unit=A 12365. 12645 /rpt_family="Alu" 13727. 13750 /note="(AC)12" /rpt_type=cadem /rpt_unit=AC 13783. 14024 /rpt_family="L1" 14175. 14470 /rpt_family="Alu" complement(14906. 15259)

primer_bind 849..867
/note="forward PCR primer"
repeat_region 875..902
/note="microsatellite Mv1389"
primer_bind
/rpt_type=tandem
/rpt_unit=ca
complement(939..960)
/note="reverse PCR primer"
BASE COUNT 420 a 449 c 383 g 310 t 101 others
ORIGIN
Query Match 16.3%; Score 24; DB 19; Length 1663;
Best Local Similarity 31.6%; Pred. No. 1.06e+00;
Matches 31; Conservative 30; Mismatches 37; Indels 0; Gaps 0;
Db 144 KGGCTGTTAGKRCACATTTTDDGMRDGRSMCWGSYMMDKTBGSKKTHSGCTTTGKG 203
OY 9 GGGTTCTTGTCTATTATTATGAGAAACCGCTCATTGTCACGCGACTGTGAGGC 68
Db 204 GKCMACRVTTGGSGTYTBTKKKCCATKGGGCGYHG 241
OY 69 CCCCACTCAGGCCACGCCCTGGCCCCCTTGTACTTG 106
RESULT 11
LOCUS HS242N11 173321 bp DNA HTG 29-SEP-1998
DEFINITION Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone
242N11; HTGS phase 1.
ACCESSION AL023655
NID 93676185
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 173321)
Westhorp,J.
Direct Submission
Submitted (29-SEP-1998) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humgen@y.sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Sep 30, 1998 this sequence version replaced g1:3618140.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known: 800 n's separate
segments.. Unfinished: dj242N11 Contig.ID: 00775 acc-AL023655
length: 1258 bp Unfinished: dj242N11 Contig.ID: 00715
acc-AL023655 length: 1075 bp Unfinished: dj242N11 Contig.ID:
01883 acc-AL023655 length: 1073 bp Unfinished: dj242N11
Contig.ID: 03048 acc-AL023655 length: 167515 bp.

*** WARNING: Phase 1 High Throughput Genome Sequence ***

* This sequence is unfinished. When sequencing is complete,
* the sequence data presented in this record will be replaced
* by a single finished sequence with the same accession number.
FEATURES
source
1..173321
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="242N11"
BASE COUNT 51256 a 32528 c 34394 g 52742 t 2401 others
ORIGIN
Query Match 16.3%; Score 24; DB 17; Length 173321;
Best Local Similarity 81.6%; Pred. No. 1.06e+00;
Matches 31; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Db 109969 TACATTATTATTATATATATTATTCTCTTC 109606

|||||
CP 143 TACATTATTATTATATATATCTGACTCGCTTC 106
RESULT 12
LOCUS MVU92534 1663 bp DNA MAM 06-APR-1997
DEFINITION Mustela vison microsatellite sequence Mv1389.
ACCESSION U92534
NID g1928977
KEYWORDS
SOURCE American mink.
ORGANISM Mustela vison
Eukaryota; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Carnivora; Fissipedidae; Mustelidae; Mustela.
1 (bases 1 to 1663)
Brusgaard,K.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE 2 (bases 1 to 1663)
AUTHORS
TITLE
JOURNAL
Direct Submission
Submitted (10-MAR-1997) Breeding and Genetics, Danish Institute of
Animal Science, Blichersalle 825, Tjele 8300, Denmark
location/Qualifiers
1..1663
/organism="Mustela vison"
/db_xref="taxon:9667"
/chromosome="8"
/map="8q1"
primer_bind 849..867
/note="forward PCR primer"
repeat_region 875..902
/note="microsatellite Mv1389"
/rpt_type=tandem
/rpt_unit=ca
complement(939..960)
/note="reverse PCR primer"
BASE COUNT 420 a 449 c 383 g 310 t 101 others
ORIGIN
Query Match 15.6%; Score 23; DB 19; Length 1663;
Best Local Similarity 27.7%; Pred. No. 3.50e+00;
Matches 23; Conservative 30; Mismatches 30; Indels 0; Gaps 0;
Db 141 ACEKGGCTGTTAGKRCACATTTTDDGMRDGRSMCWGSYMMDKTBGSKKTHSGCTTTB 200
CP 116 ACTGCTGTTCCAGTACCAAGGGGGCCAGGCTGCTGAGTGGGCGCTCACAGTGC 57
Db 201 GKCGKMACRVTTGGSGGTYTBT 223
CP 56 GCTGGACAAAGTGACCGGTTTT 34
RESULT 13
LOCUS AC005367 81741 bp DNA PRI 01-AUG-1998
DEFINITION Homo sapiens chromosome 5, PAC clone 156M17 (LENL H150), complete
sequence.
ACCESSION AC005367
NID 93367507
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 81741)
Church,D.M., Yang,J., Bocian,M., Shiang,R. and Wasmuth,J.J.
A high-resolution physical and transcript map of the Cri du chat
region of human chromosome 5p
Genome Res. 7 (8), 787-801 (1997)
JOURNAL
MEDLINE
97413160
REFERENCE 2 (bases 1 to 81741)
AUTHORS
Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,

repeat_region complement(80527..80869)
 repeat_region /rpt_family="THE1"
 repeat_region complement(81164..81423)
 repeat_region /rpt_family="ALU"
 BASE COUNT 24685 a 18143 c 15743 g 23170 t
 ORIGIN

Query Match 15.6% Score 23: DB 26: Length 81741:
 Best Local Similarity 81.1% Pred. No. 3.50e+00;
 Matches 30: Conservative 0: Mismatches 7: Indels 0: Gaps 0:

Db 38495 ATATCATCTGCAATTGGCTCAATTCACCAAGTGG 38531
 Cp 128 ATATATCTGTAATCTGGTTCACAGTACCAAGGGG 92

RESULT 14 AC003089 126198 bp DNA PRI 06-NOV-1997
 LOCUS Human BAC clone RG180F08A, complete sequence.
 DEFINITION AC003089
 ACCESSION 92588614
 NID
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 AUTHORS Wilson, R., Smith, A., Elliott, G., Kramer, J., Latreille, P. and Keppler, D.
 TITLE The sequence of H. sapiens BAC clone RG180F08A
 JOURNAL Unpublished (1997)
 REFERENCE 2 (bases 1 to 126198)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (06-NOV-1997) Department of Genetics, Washington
 COMMENT University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 SUBMITTED BY: Genome Sequencing Center
 Department of Genetics
 Washington University
 St. Louis MO 63108, USA
 mailto://genome.wustl.edu/gsc
 mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (http://www.resgen.com).
 VECTOR: pBelobAC11
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The actual start of this clone is at base position 1 of RG180F08A; the actual end is unknown. The sequence ends at base 126198. The orientation of this clone is unknown. This sequence was a contaminating clone in the database of RG180F08, the real name of the clone is unknown.

This clone contains single stranded regions from 18047-18064, 31477-31498, 41155-41228, 55004-55008, 63371-63414, 106792-106803,

FEATURES and 123800-125250 which is in a tandem repeat region.
 source Location/Qualifiers

1..126198
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="UL"
 /clone="RG180F08A"
 /clone_11b="C17B-978SK-B"
 /map="unknown"
 85..1478
 /rpt_family="L1"
 repeat_region complement(2169..2535)
 /rpt_family="MER"
 8077..8152
 /rpt_family="L1"
 repeat_region 11140..11187
 /rpt_family="L1"
 11492..11512
 /rpt_family="L1"
 repeat_region complement(12283..12409)
 /note="match to EST T0654 (NID:9317803)"
 gene /note="match to EST T0654 (>49658)"
 /gene="WUGSC:H.RG180F08A.1"
 /gene="WUGSC:H.RG180F08A.1"
 /note="match to U96136 (PID:92073006)"
 /product="delta-catenin"
 /db_xref="PID:92588615"
 /translation="REFGWDPELVIGMIOHPPSVQSNAAVYLOHLCFGDNKTKA
 EIRROGGLIYDLDRHRTVEHRSAGLRLNLYIGKANDDKRIALKNCGGIPALVRL
 LKRTIDLEIREL"
 repeat_region complement(13523..13814)
 /rpt_family="ALU"
 14213..14249
 /rpt_family="L1"
 repeat_region complement(15648..15940)
 /rpt_family="ALU"
 complement(16266..16286)
 /rpt_family="L1"
 repeat_region complement(16287..16510)
 /rpt_family="ALU"
 18009..18063
 /rpt_family="L1"
 repeat_region 18273..18319
 /rpt_family="L1"
 19848..20219
 /rpt_family="L1"
 21456..21532
 /rpt_family="L1"
 repeat_region 22020..22224
 /rpt_family="L1"
 22519..22764
 /rpt_family="ALU"
 22960..23243
 /rpt_family="ALU"
 24096..24121
 /rpt_family="L1"
 25710..25744
 /rpt_family="L1"
 33190..33215
 /rpt_family="L1"
 35360..35647
 /rpt_family="ALU"
 37492..37525
 /rpt_family="L1"
 39255..39363
 /rpt_family="ALU"
 40895..41178
 /rpt_family="ALU"
 41199..41226
 /rpt_family="L1"
 repeat_region complement(41330..41615)

```

repeat_region /rpt_family="ALU" 42197. .42368
repeat_region /rpt_family="ALU" 42902. .43216
repeat_region /rpt_family="L1" 44103. .45634
repeat_region /rpt_family="THR" complement(45635. .45820)
repeat_region /rpt_family="ALU" complement(45865. .45997)
repeat_region /rpt_family="ALU" 48638. .48689
repeat_region /rpt_family="L1" complement(49861. .50151)
repeat_region /rpt_family="ALU" complement(50342. .50495)
repeat_region /rpt_family="ALU" 53802. .53828
repeat_region /rpt_family="L1" complement(57143. .57432)
repeat_region /rpt_family="ALU" 62626. .62661
repeat_region /rpt_family="L1" complement(63381. .63671)
repeat_region /rpt_family="ALU" complement(66167. .66341)
repeat_region /rpt_family="ALU" complement(66417. .66524)
repeat_region /rpt_family="ALU" 70316. .70558
repeat_region /rpt_family="ALU" 75313. .75603
repeat_region /rpt_family="ALU" 76262. .76289
repeat_region /rpt_family="L1" 77475. .77767
repeat_region /rpt_family="ALU" complement(78542. .78837)
repeat_region /rpt_family="ALU" complement(79715. .80006)
repeat_region /rpt_family="ALU" 81129. .81427
repeat_region /rpt_family="ALU" 83300. .83317
repeat_region /rpt_family="L1" complement(90357. .90645)
repeat_region /rpt_family="ALU" 96209. .96498
repeat_region /rpt_family="ALU" complement(97108. .97628)
repeat_region /rpt_family="L1" 99857. .99940
repeat_region /rpt_family="L1" complement(104120. .104208)
repeat_region /rpt_family="L1" complement(105301. .105674)
repeat_region /rpt_family="ALU" 106667. .106698
repeat_region /rpt_family="L1" complement(106702. .106726)
repeat_region /rpt_family="L1" 107454. .109911
repeat_region /rpt_family="L1" complement(108498. .108916)
repeat_region /rpt_family="L1" complement(111096. .111380)
repeat_region /rpt_family="ALU" 111381. .111518
misc_feature /note="match to EST AA426571 (NID:92107392) zv83a05.s1"
misc_feature complement(111564. .111993)
misc_feature /note="match to EST AA425097 (NID:92107548) zv83a05.r1"

```

Note: remainder of annotations omitted.

```

Query Match 15.6% Score 23; DB 26; Length 126198;
Best Local Similarity 81.1% Pred. No. 3,50e+00;
Matches 30; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 84541 ATATCATCTGCATATTTGGCTCCAAATACCAAGTGG 84577
CP 128 ATATATCTGTACTTCGGTCCAGTACCAAGGGG 92

RESULT 15
LOCUS GBV84940 183 bp RNA VRL 21-MAR-1997
DEFINITION GBV-A-like virus recovered from Saginus nigricollis 96. 5'
ACCESSION U84940
NID 91890719
KEYWORDS
SOURCE GBV-A-like virus.
ORGANISM GBV-A-like virus.
REFERENCE 1 (bases 1 to 183)
AUTHORS Bukh, J. and Apgar, C. L.
TITLE Five new or recently discovered (GBV-A) virus species are
of the Flaviviridae
JOURNAL Virology 229 (2), 429-436 (1997)
MEDLINE 97271304
REFERENCE 2 (bases 1 to 183)
AUTHORS Bukh, J. and Apgar, C. L.
TITLE Direct Submission
JOURNAL Submitted (13-JAN-1997) Hepatitis Viruses Section, Laboratory of
Infectious Diseases, National Institutes of Health, Building 7, Room 201, 7
Center Dr MSC 0740, Bethesda, MD 20892-0740, USA

FEATURES
source 1.183
/organism="GBV-A-like virus"
/strain="SN 96"
/specific_host="Saginus nigricollis"
/note="corresponds to nts. 329-511 of GBV-A (5' noncoding
region)"
/db_xref="taxon:56547"

BASE COUNT 39 a 58 c 60 g 26 t
ORIGIN

Query Match 15.0% Score 22; DB 33; Length 183;
Best Local Similarity 92.3% Pred. No. 1,13e+01;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 82 CCAAGGGGGGGCCGGCGCTGCCTGA 107
CP 100 CCAAGGGGGGGCCAGCGCTGCCTGA 75

```

Search completed: Tue Dec 22 04:12:14 1998
 Job time : 2/4 secs.

Osaka University
3-1, Yamadaoka
Suita, Osaka, 565
Japan
Phone: 06-877-5111
Fax : 06-875-1922.

FEATURES

Source

1. 304
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HL60"
/cell_type="promyelocyte"
/dev_stage="adult"
/sex="female"

BASE COUNT 84 a 41 c 67 g 106 t 6 others

ORIGIN

Query Match 100.0%; Score 298; DB 7; Length 304;

Best Local Similarity 100.0%; Pred. No. 0.00e+00;

Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GATCTAAGCAGTNGTGTGTTACATGTTTCTACACATTATCCTTAAAGTTGT 60
|||||
Qy 1 GATCTAAGCAGTNGTGTGTTACATGTTTCTACACATTATCCTTAAAGTTGT 60
Db 61 TGAGAGAGTGTATTACCTCCCAAGTTGGAAGCAGGGAATTCCTCCAGTCTTA 120
|||||
Qy 61 TGAGAGAGTGTATTACCTCCCAAGTTGGAAGCAGGGAATTCCTCCAGTCTTA 120
Db 121 GTTTTCACACAGAGATATGTAGTAGCAAGATTTGCTGCTTACATATAGTGT 180
|||||
Qy 121 GTTTTCACACAGAGATATGTAGTAGCAAGATTTGCTGCTTACATATAGTGT 180
Db 181 ATGTATGATATATGTAATNGTGTGTTAAAGCCTGATCTATTTTTCATANGCAATG 240
|||||
Qy 181 ATGTATGATATATGTAATNGTGTGTTAAAGCCTGATCTATTTTTCATANGCAATG 240
Db 241 TTAAGCAAGAGCCTCCCTGATTTGANGAGCAGGTTTTCATATATGATTTTNGGA 300
|||||
Qy 241 TTAAGCAAGAGCCTCCCTGATTTGANGAGCAGGTTTTCATATATGATTTTNGGA 300
Db 301 TAAA 304
|||||
Qy 301 TAAA 304

RESULT

LOCUS AA463716 387 bp mRNA EST 10-JUN-1997

DEFINITION aa07d11.s1 Soares Nhmpu S1 Homo sapiens cDNA clone 812565 3', mRNA

ACCESSION AA463716

NTD 92188600

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae;

Homo.

1 (bases 1 to 387)

REFERENCE AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Scheilenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.

Washu-Merck EST Project 1997

Unpublished (1997)

TITLE

Washu-Merck EST Project 1997

Unpublished (1997)

CONTACT: Wilson RK

Washu-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: estewatson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: 41m3 fwd. ET from Amersham High quality sequence stop: 345.

FEATURES

Source

1. 387
/organism="Homo sapiens"
/note="Organ: mixed (see below); Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbH, pregnant uterus MbHPU, and fetal heart MbH19M) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

/db_xref="taxon:9606"
/clone="812565"
/clone_lib="Soares Nhmpu S1"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
complement(<1. >387)
/db_xref="GDB:6043475"

BASE COUNT

146 a 77 c 52 g 112 t

ORIGIN

Query Match 97.7%; Score 291; DB 7; Length 387;

Best Local Similarity 97.7%; Pred. No. 0.00e+00;

Matches 297; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Db 15 TTTATCCCAAAAATCATATAAATGAAAACTGCTTCAATGACGAGGAGCCTTGGC 74
|||||
Cp 304 TTTATCCCAAAAATCATATAAATGAAAACTGCTTCAATGACGAGGAGCCTTGGC 245
Db 75 T-AACATGTCATATGAAATCAGATCAGCTTTACACACATTTACATATATAT 133
|||||
Cp 244 TTTACATGTCATATGAAATCAGATCAGCTTTACACACATTTACATATATAT 185
Db 134 ACATACACATATATGAGCAGCAATACCTTGTCTACTTACATATCTCTGTGGA 193
|||||
Cp 184 ACATACACATATATGAGCAGCAATACCTTGTCTACTTACATATCTCTGTGGA 125
Db 194 AAACTAGGACACTGGGAATTCCTGCTTCCAACTTGGAGGTAATACACCTCT 253
|||||
Cp 124 AAACTAGGACACTGGGAATTCCTGCTTCCAACTTGGAGGTAATACACCTCT 65
Db 254 CTCACCACTTTTAAAGATGAATGTGTAACATGTAACACACACACCTCTTA 313
|||||
Cp 64 CTCACCACTTTTAAAGATGAATGTGTAACATGTAACACACACACCTCTTA 5
Db 314 GATC 317
|||||
Cp 4 GATC 1

RESULT

LOCUS AA609872 421 bp mRNA EST 02-MAR-1998

DEFINITION af08c06.s1 Soares testis NHT Homo sapiens cDNA clone 1031050 3',

sequence.

ACCESSION AA609872

NTD 92458300

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 421)

REFERENCE AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Scheilenberg,K., Steptoe,M., Tan,F.,

TITLE Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
JOURNAL WashU-NCI human EST Project
COMMENT Unpublished (1997)

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu

This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (infoimage.lnl.gov) for further information.
Insert Length: 1117 Std Error: 0.00
Seq primer: -40m13 fwd. EF from Amersham
High quality sequence stop: 351.

FEATURES
source

1. 421
/organism="Homo sapiens"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer 15'.
TGTACCAATCTGAAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 371.
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_1lb="1031050"
/sex="male"
/lab_host="DH10B"
/lab_host="55 g 126 t

BASE COUNT 160 a 160 c 55 g 126 t
ORIGIN

Query Match 97.7%; Score 291; DB 13; Length 421;

Best Local Similarity 97.7%; Pred. No. 0.00e+00; Mismatches 6; Indels 1; Gaps 1;

Matches 297; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Db 26 TTATCCCAAAATACATTAATGAAGAAACCTGCTTCAATAGCAGGAGGCTTGGC 85
Cp 304 TTATCCCAAAATACATTAATGAAGAAACCTGCTTCAATAGCAGGAGGCTTGGC 245
Db 86 T-AACATTGTCAATGAAATCAGTATCAGCTCTTAAACACAAATTACATATATACAT 144
Cp 244 TTAACATTGNCATATGAAATCAGTATCAGCTCTTAAACACAAATTACATATATACAT 185
Db 145 ACATACACACTATATATAGCAGCAAAATCTTGTCTACTTACACATATTCCTGCTGGA 204
Cp 184 ACATACACACTATATATAGCAGCAAAATCTTGTCTACTTACACATATTCCTGCTGGA 125
Db 205 AACAGTAGACACTGGGAATTCCTGCTTCCACCTTGGGAAGTAATPACAACTCT 264
Cp 124 AACAGTAGACACTGGGAATTCCTGCTTCCACCTTGGGAAGTAATPACAACTCT 65
Db 265 CTCACCACTTTTAAAGATGAAATGTGTAGAAACATGTAAACACAACTGCTTTA 324
Cp 64 CTCACCACTTTTAAAGATGAAATGTGTAGAAACATGTAAACACAACTGCTTTA 5
Db 325 GATC 328
Cp 4 GATC 1

RESULT 4
LOCUS AA496258 377 bp mRNA EST 18-ANG-1997
DEFINITION aa24b10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:814171 3',
ACCESSION AA496258
NID 92229579
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae;
Homo.

REFERENCE 1 (bases 1 to 377)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997)

JOURNAL
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LNLN at:

www.bio.lnl.gov/db/ftp/image/image.html

Seq primer: -40m13 fwd. EF from Amersham
High quality sequence stop: 351.

Location/Qualifiers

FEATURES
source

1. 377
/organism="Homo sapiens"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTACCAATCTGAAGTGGAGCGCGCCCAATTTTCTTTTCTTTT-
371]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_1lb="IMAGE:814171"
/clone_1lb="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/db_xref="GDB:6032124"

BASE COUNT 117 a 117 c 44 g 146 t
ORIGIN

Query Match 97.0%; Score 289; DB 8; Length 377;

Best Local Similarity 97.4%; Pred. No. 0.00e+00; Mismatches 7; Indels 1; Gaps 1;

Matches 296; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

Db 74 TTATCCCAAAATACATTAATGAAGAAACCTGCTTCAATAGCAGGAGGCTTGGC 133
Cp 304 TTATCCCAAAATACATTAATGAAGAAACCTGCTTCAATAGCAGGAGGCTTGGC 245
Db 134 T-AACATTGTCAATGAAATCAGTATCAGCTCTTAAACACAAATTACATATATACAT 192
Cp 244 TTAACATTGNCATATGAAATCAGTATCAGCTCTTAAACACAAATTACATATATACAT 185
Db 193 ACATACACACTATATATAGCAGCAAAATCTTGTCTACTTACACATATTCCTGCTGGA 252
Cp 184 ACATACACACTATATATAGCAGCAAAATCTTGTCTACTTACACATATTCCTGCTGGA 125
Db 253 AACAGTAGACACTGGGAATTCCTGCTTCCACCTTGGGAAGTAATPACAACTCT 312
Cp 124 AACAGTAGACACTGGGAATTCCTGCTTCCACCTTGGGAAGTAATPACAACTCT 65
Db 313 CTCACCACTTTTAAAGATGAAATGTGTAGAAACATGTAAACACAACTGCTTTA 372

|||||
Cp 64 CTCACAACTTTTAAAGATGAAATGTAGAAACATGTAACAACACACACTGCTTTA 5
Db 373 GATC 376
|||||
Cp 4 GATC 1

RESULT 5
LOCUS AA760717 474 bp mRNA EST 07-FEB-1998
DEFINITION n213d07.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1287661,
mRNA sequence.
ACCESSION AA760717
NID 92809647
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 474)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html
Insert Length: 2006 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 447.
Location/Qualifiers
1. 474
/organism="Homo sapiens"
/note="Vector: p773p-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTCCATCTCAAGTGGAGGCGCCGCTCATTTTCTTTTCTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_image="1287661"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/lab_host="DH10B"

BASE COUNT 174 a 85 c 61 g 154 t
ORIGIN

Query Match 88.6%; Score 264; DB 12; Length 474;
Best Local Similarity 94.5%; Pred. No. 0.00e+00;
Matches 299; Conservative 0; Mismatches 14; Indels 3; Gaps 3;

Db 52 TTATCCCAAAATCATTAATAATGAAACCTGCCCCCTTCAATGCGAGGAGGCGCTTTG 111
|||||
Cp 304 TTATCCCAAAATCATTAATAATGAAACCTGCCCCCTTCAATGCGAGGAGGCGCTTTG 247
|||||

Db 112 CCTAACAATTCATATGAAAAAGATATCAGCCCTTTAACAACACAAATTTACATATATA 171
|||||
Cp 246 CCTTACACTTCATATGAAAAATCGATACAGC-TCCTTACACACAAATTTACATATATA 188
|||||
Db 172 CATACATACACATATATATAGACAGCAATATCTTGCTACTTACACATATCTCTGCT 231
|||||
Cp 187 CATACATACACATATATATAGACAGCAATATCTTGCTACTTACACATATCTCTGCT 128
|||||
Db 232 GGAAGAACTAGACACTGGGAATTCCTGCTTCCAACTTTGGAGAGTAAATCAAC 291
|||||
Cp 127 GGAAGAACTAGACACTGGGAATTCCTGCTTCCAACTTTGGAGAGTAAATCAAC 68
|||||
Db 292 TCTCTCAACAATTTTAAAGATGAATGTAGAAACATGTAAACAACCAACCTGCT 351
|||||
Cp 67 TCTCTCAACAATTTTAAAGATGAATGTAGAAACATGTAAACAACCAACCTGCT 8
|||||
Db 352 TTAGATC 358
|||||
Cp 7 TTAGATC 1

RESULT 6
LOCUS SSC11609 443 bp RNA EST 29-OCT-1997
DEFINITION Sus scrofa mRNA; expressed sequence tag (3'; clone c11909), mRNA
sequence.
ACCESSION F22894
NID 92578461
KEYWORDS EST; expressed sequence tag.
SOURCE Pig.
ORGANISM Sus scrofa
Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Artiodactyla; Suiformes; Suidae; Sus.
REFERENCE 1 (bases 1 to 443)
Winteroe, A.K. and Fredholm, M.
Evaluation and characterization of a porcine small intestine cDNA
library
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 443)
Winteroe, A.K.
Direct Submission
Submitted (26-OCT-1997) Winteroe A.K., The Royal Veterinary and
Agricultural University, Department of Animal Science and Animal
Health, Division of Animal Genetics, Bulowsevej 13, 1870
Frederiksberg C, DENMARK
Location/Qualifiers
1. 443
/organism="Sus scrofa"
/db_xref="taxon:9823"
/tissue_type="small intestine"
/clone_lib="directionally cloned cDNA in XLI-blue MRF"
1. >443
/note="expressed sequence tag"

BASE COUNT 156 a 93 c 61 g 132 t 1 others
ORIGIN

Query Match 30.2%; Score 90; DB 10; Length 443;
Best Local Similarity 81.8%; Pred. No. 5.66e-103;
Matches 135; Conservative 0; Mismatches 27; Indels 3; Gaps 3;

Db 186 ACACACACACACACACACATATAGTAGTATGACCTTTCTACTTACACATA 245
|||||
Cp 196 ACATATATACATATATACACACATATATGACACCAAT-ACCTTGCTACTTACACATA 138
|||||
Db 246 TTCTCTGTAGAGATTAGAGAGTGGCAATTCCTGCTCTTCCAGCCTTGGAGAG 305
|||||
Cp 137 TTCCTCTGTGGAAGACAGACACTGGGAATTCCTGCTTCCAGCCTTGGAGAG 79
|||||
Db 306 TAAATACACCTCTCTGAACGCTTTTAAAGATGAACAATGT 350
|||||
Cp 78 TAAATACACCTCTCTGAACGCTTTTAAAG-GATGAAGATGT 35
|||||


```

RESULT 7 AA754459 252 bp mRNA EST 20-JAN-1998
LOCUS 97SN1787 Rice Immature Seed Lambda ZAPII CDNA Library Oryza sativa
DEFINITION CDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID 92801165
KEYWORDS EST.
SOURCE rice.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Charophyta; Embryophyta group;
Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;
Poales; Poaceae; Oryza.
1 (bases 1 to 252)
REFERENCE 1
AUTHORS Nahn,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Lee,M.C. and Eun,M.Y.
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
Lee,M.C. and Eun,M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL Unpublished (1998)
COMMENT
Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sunn20.asi.re.kr
Submitted by Baek Hie Nahn, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bhahn@bioserver.myongji.ac.kr
Seq primer: M13 Reverse primer.
FEATURES
source
1..252
Location/Qualifiers
/organism="Oryza sativa"
/cultivar="Milyang23"
/notes="Vector: pBluescript SK(+); site_1: EcoRI; site_2:
XhoI; Directional CDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII CDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT 5 a 21 c 12 g 35 t 179 others
ORIGIN
Query Match 17.8%; Score 53; DB 12; Length 252;
Best Local Similarity 11.0%; Pred. No. 2,94e-44;
Matches 23; Conservative 108; Mismatches 74; Indels 4; Gaps 4;
Db 29 BWAVCAVSHGNYSVNHCIBRGTHCDCKNNVMSITMGIVNMBVSGDMHYBVBNTKVD 88
Cp 300 TCCCNAAAATGATATTAAGANANACCTGCTCNCATGACGAGCGCTTTCCTTAA 241
Db 89 VGNHTRCRMBRYTBMAYHDYTNCBYNNNDYHMBMBY-BBTGCMTCMM-CMBHYN 146
Cp 240 C-ATGNCATATGAAATTCATACGCTCTTAAACACACATTTACATATATACATACA 182
Db 147 TKCTASGWHSTSYNDVKSSTNTGVTBSYDKSMHGYSBVBKXHYKVTTRATSYTCV 206
Cp 181 TACACACTATATATGTAACGACACAATACTTGTCTACTTACACATATTCCTCGTGGA 122
Db 207 RKTCVM-WMTKKVKKHVVBBGCHTDS 234
Cp 121 CTAGACACTGGGAAATTCCTCCTTC 93
RESULT 8 AA273441 354 bp mRNA EST 28-MAR-1997
LOCUS vc01c11.r1 Soares mouse lymph node NbMLN Mus musculus CDNA clone
DEFINITION 765236 5', mRNA sequence.
ACCESSION AA273441
NID 91913436
KEYWORDS EST.

```

```

SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
REFERENCE 1 (bases 1 to 354)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kneba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Yan,E., Underwood,K., Moore,B.,
Theisling,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:466156
Seq primer: -28m13 rev2 ET from Amerisham.
FEATURES
source
1..354
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/notes="Vector: pT7R3D-Pac (Pharmacia) with a modified
polylinker; site_1: Not I; site_2: Eco RI; 15'
TCTTACCATCTGAGAGTGGAGCGCGCGGATCTTTTCTTTTCTTTTCTTTTCTTTT
3'"); double-stranded CDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library constructed and
normalized by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:10090"
/clone="765236"
/clone_lib="Soares mouse lymph node NbMLN"
/sex="male"
/dev_stage="4 weeks"
/lab_host="DH10B"
BASE COUNT 87 a 69 c 85 g 113 t
ORIGIN
Query Match 16.1%; Score 48; DB 24; Length 354;
Best Local Similarity 86.9%; Pred. No. 8.08e-37;
Matches 53; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Db 291 AGCCAAAGCCTCTGCAATTTGAGACAGAGTTTCATTTATGATTTTGACATTA 350
Cp 244 AGCCAAAGCCTCTGCAATTTGAGACAGAGTTTCATTTATGATTTTGACATTA 303
Db 351 A 351
Cp 304 A 304
RESULT 9 AA955078 430 bp mRNA EST 07-MAY-1998
LOCUS UI-R-A1-ey-h-11-0-UI.s1 UI-R-A1 Rattus norvegicus CDNA clone
DEFINITION UI-R-A1-ey-h-11-0-UI 3', similar to g11913436 (p1AA273441)AA273441
vc01c11.r1 Soares mouse lymph node NbMLN Mus musculus CDNA clone
765236 5', mRNA sequence.
ACCESSION AA955078
NID 93119127
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

```

REFERENCE 1 (bases 1 to 430)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT

Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence tag present in the cDNA between the NotI site and the oligo-dT track (not shown) served to identify it as a clone from the normalized adult lung library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics
 Seq primer: M13 Forward.

FEATURES
 source

1.430
 Location/Qualifiers
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The UR-R-A1 library is a subtracted library derived from the UR-R-A0 library. The UR-R-A0 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UR-R-A1) was constructed as follows: PCR amplified cDNA inserts from a pool of approximately 3,840 UR-R-A0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UR-R-A0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UR-R-A1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996).
 /db_xref="taxon:10116"
 /clone="UR-R-A1-ey-h-11-0-UR"
 /clone_lib="UR-R-A1"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"

BASE COUNT 155 a 95 c 79 g 101 t
 ORIGIN

Query Match 16.1%; Score 48; DB 14; Length 430;
 Best Local Similarity 86.9%; Pred. No. 8.08e-37;
 Matches 53; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 40 TTAATCTCAAAATACATTAATGAAGAAACCTGCTCCAAATGACAGACAGCGCTTGCC 99
 |||||
 Cp 304 TTATCCCAAAATACATTAATGAAGAAACCTGCTCCAAATGACAGAGCGCTTGCC 245
 Db 100 T 100
 Cp 244 T 244

RESULT 10 A1008917 449 bp mRNA EST 15-JUN-1998
 LOCUS EST03368 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
 DEFINITION REMB37 3' end, mRNA sequence.
 ACCESSION A1008917
 NID 93222749

KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 REFERENCE 1 (bases 1 to 449)
 AUTHORS Lee,N.H., Glodex,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.
 TITLE Rat Genome Project: Generation of a Rat EST (REEST) Catalog & Rat Gene Index
 JOURNAL Unpublished (1998)
 COMMENT

Contact: Lee, NH
 ATCC
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.

FEATURES
 source

1.449
 Location/Qualifiers
 /organism="Rattus sp."
 /note="Vector: pT73Pac; Site.1: EcoRI; Site.2: NotI"
 /db_xref="taxon:10118"
 /clone="REMB37"
 /clone_lib="Normalized rat embryo, Bento Soares"
 /dev_stage="embryo 8, 12, 18 dpc"
 BASE COUNT 159 a 101 c 93 g 96 t
 ORIGIN

Query Match 15.8%; Score 47; DB 15; Length 449;
 Best Local Similarity 87.9%; Pred. No. 2.36e-35;
 Matches 51; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 2 ATCTCAAAATACATTAATGAAGAAACCTGCTCCAAATGACAGACAGCGCTTGCT 59
 |||||
 Cp 301 ATCCCAAAATACATTAATGAAGAAACCTGCTCCAAATGACAGAGCGCTTGCT 244

RESULT 11
 LOCUS AA754459 252 bp mRNA EST 20-JAN-1998
 DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
 cDNA clone 97SN1787, mRNA sequence.
 ACCESSION AA754459
 NID 92801165
 KEYWORDS EST.
 SOURCE
 ORGANISM Oryza sativa
 rice.
 Eukaryotes; Viridiplantae; Charophyta/Embryophyta group; Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
 REFERENCE 1 (bases 1 to 252)
 AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.
 TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed Unpublished (1998)
 COMMENT

CONTACT: Eun M.Y.
 Department of Cytogenetics
 National Inst. of Agri. Sci. and Tech, RDA
 Suwon, Kyungido, Korea
 Tel: 82 331 290 0301
 Fax: 82 331 290 0307
 Email: myeunsun20.asi.re.kr
 Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongsin, Korea. 449-728 dnamedioserver.myongji.ac.kr
 Seq primer: M13 Reverse Primer.
 Location/Qualifiers
 1.252
 /organism="Oryza sativa"
 /culturvar="Milyang23"

Matches 15: Conservative 80: Mismatches 64: Indels 2: Gaps 2:

Db 79 TBYISWNVDTGTTGKTVNHNHSGMNNRCSNVYVWBYAYCDYBHDRAHVDT 138
 Cp 238 TTGNCATATGAAATACAGTATCAGCTCTTTAAACACANATTACATATATACATACATAC 179

Db 139 RCTNDRCGNCNTASDNGTSATKRTVGTGKTDSDCGGCGW-RKVTYGS-SBYBRGCVNMY 196
 Cp 178 ACACATATATGATACACAGCAATCTGCTTACTTACACATATCTCTCTGCGGAAACTA 119

Db 197 RTTSMWTDKSTKMSMDMSRSTRVHYGMBNKRKGRSMRN 237
 Cp 118 GGACACTGGAAATTCCTCCCTCTTCAACCTTGGGAAGT 78

RESULT 14
 LOCUS AA730318 363 bp mRNA EST 27-JUN-1998
 DEFINITION mw41101.s1 NCI_CGAP_Ew1 Homo sapiens CDNA clone IMAGE:1249177
 similar to contains Alu repetitive element; contains element MER10
 repetitive element ; , mRNA sequence.

ACCESSION AA730318
 NID 92751522
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 363)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
 Ph.D.
 cDNA Library Preparation: David B. Krizman, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Insert Length: 744 Std Error: 0.00
 Seq primer: -40m3 fwd. Et from Amersham
 High quality sequence stop: 363.
 Location/Qualifiers
 1..363
 /organism="Homo sapiens"
 /note="Vector: PAMPI0: mRNA made from Ewing's sarcoma,
 cDNA made by oligo-dT priming. Non-directionally cloned.
 Size-selected on agarose gel, average insert size 600 bp.
 Reference: Krizman et al. (1996) Cancer Research
 56:5380-5383."
 /db_xref="taxon:9606"
 /clone="IMAGE:1249177"
 /clone_id="NCI_CGAP_Ew1"
 /tissue_type="Ewing's sarcoma"
 /lab_host="DH10B"
 /lab_host="DH10B"

BASE COUNT 110 a 65 c 67 g 121 t

ORIGIN

Query Match 9.4%: Score 28; DB 12; Length 363;
 Best Local Similarity 70.3%: Pred. No. 1.31e-09;
 Matches 45: Conservative 0: Mismatches 19: Indels 0: Gaps 0:

Db 149 TTGAATATGACACATTTTAAATGCTTTATGACATTTACATATATACATACATAG 208
 Cp 238 TTGNCATATGAAATACAGTATCAGCTCTTTAAACACANATTACATATATACATACATAC 179
 Db 209 ATAC 212

Cp 178 ACAC 175

RESULT 15
 LOCUS H10802 485 bp mRNA EST 26-JUN-1995
 DEFINITION ym04a11.r1 Homo sapiens CDNA clone 46650 5' similar to contains
 MRR32 repetitive element ; .
 ACCESSION H10802
 NID 9875622
 KEYWORDS EST.
 SOURCE human clone-46650 library-Soares infant brain 1MB vector-Lafmid BA
 host-DH10B (ampicillin resistant) primer-M13rev Rsite1-Not I
 Rsite2-Hind III whole brain from a 73 days post natal female. 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5',
 AACTGGAAGAATTCGGCGCCGACAGAAATTTTCTTTTCTTTT 3']; double-stranded
 cDNA was ligated to Hind III adaptors (Pharmacia), digested with
 Not I and directionally cloned into the Not I and Hind III sites of
 the Lafmid BA vector. Library went through one round of
 normalization. Library constructed by Bento Soares and M.Fatima
 Bonaldo.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata;
 Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
 Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
 Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 485)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P., and
 Wilson, R.
 The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT

TITLE
 JOURNAL
 COMMENT

GDB: 600-419-191
 Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 High quality sequence stops: 335
 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Location/Qualifiers
 1..485
 /organism="Homo sapiens"
 /clone="46650"

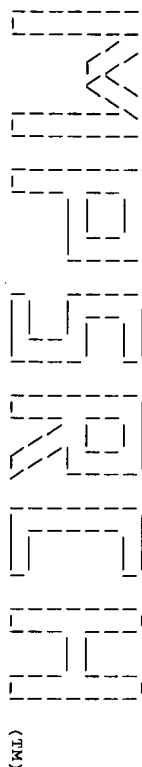
BASE COUNT 120 a 84 c 94 g 163 t 24 others

ORIGIN

Query Match 9.1%: Score 27; DB 16; Length 485;
 Best Local Similarity 93.5%: Pred. No. 2.04e-08;
 Matches 29: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

Db 102 ATATAGTATATATATATATATATATATATATATAT 132
 Oy 170 ATATAGTGTATATATATATATATATATATATATAT 200

Search completed: Tue Dec 22 04:35:03 1998
 Job time : 476 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_nm n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 04:35:23 1998: Maspar time 61.47 Seconds

Tabular output not generated. 672.834 Million cell updates/sec

Title: >US-08-530-112A-1017
Description: (1-304) from US08530112A.seq
Perfect Score: 298
N.A. Sequence: 1 GATCTAAGCAGCGTNGTGT...ATATGATTTTGGGATATA 304
Comp: CTAGATTTCGTCANCAACA.....TATACATTAANCCCTATT

Scoring table: TABLE default

Nmatch STD : Dbase 0: Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 7.820; Variance 5.062; scale 1.545

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	298	100.0	304	19	T20017 Human gene signature	2,92e-173
2	40	13.4	204	1	N81164 Base substituted E.co	9.57e-10
3	37	12.4	91	9	Q51746 Oligonucleotide probe	4.23e-08
4	35	11.7	204	1	N81164 Base substituted E.co	5.09e-07
5	33	11.1	91	9	Q51746 Oligonucleotide probe	5.91e-06
6	26	8.7	67	24	T14322 Primer used in the la	2.25e-02
7	25	8.4	68	32	T63255 Messenger RNA primer	6.93e-02
8	25	8.4	68	34	T73397 Oligonucleotide tag c	6.93e-02
9	25	8.4	70	24	T14325 Conjugate formed by 1	6.93e-02
10	25	8.4	19124	32	T72882 Plasmodium var-7 gene	6.93e-02
11	24	8.1	39	24	T14323 Sequence used in the	2.10e-01
12	23	7.7	114	12	O70467 Generic DNA sequence	6.24e-01
13	23	7.7	114	12	O70466 Generic DNA sequence	6.24e-01

Result ID	Score	Query Match	Length	DB ID	Description	Pred. No.
14	23	7.7	304	19	T20017 Human gene signature	6.24e-01
15	23	7.7	1380	30	T43928 Sequence around marke	6.24e-01
16	23	7.7	1380	30	T72048 Sequence around marke	6.24e-01
17	22	7.4	114	12	O70465 Generic DNA sequence	1.82e+00
18	22	7.4	114	12	O70465 Generic DNA sequence	1.82e+00
19	22	7.4	114	12	O70468 Generic DNA sequence	1.82e+00
20	22	7.4	154	39	V13060 Interleukin 6 non-ris	1.82e+00
21	22	7.4	501	3	N50027 Sequence encoding new	1.82e+00
22	22	7.4	2875	29	T47713 Mouse bone morphogene	1.82e+00
23	22	7.4	5511	24	T41853 CDNA encoding Plasmod	1.82e+00
24	22	7.4	9789	24	T41852 CDNA encoding Plasmod	1.82e+00
25	22	7.4	15144	29	T47715 Mouse bone morphogene	1.82e+00
26	22	7.4	133894	17	T13635 ACNPV genomic DNA clo	1.82e+00
27	21	7.0	56	26	T63872 Human H4/b gene (for	5.20e+00
28	21	7.0	56	11	O69410 Human H4/b gene for H	5.20e+00
29	21	7.0	59	5	O33502 Sequence of microsat	5.20e+00
30	21	7.0	149	15	O87012 Subtelomeric CDNA clo	5.20e+00
31	21	7.0	149	37	T96339 Subtelomeric CDNA clo	5.20e+00
32	21	7.0	149	9	O52430 Human gene signature	5.20e+00
33	21	7.0	151	19	T20274 Human gene signature	5.20e+00
34	21	7.0	305	19	T21318 Human gene signature	5.20e+00
35	21	7.0	330	8	O60584 Human brain Expressed	5.20e+00
36	21	7.0	501	3	N50023 Sequence encoding new	5.20e+00
37	21	7.0	518	10	O55238 Grapevine ribosomal c	5.20e+00
38	21	7.0	1215	31	T43468 Human gene signature	5.20e+00
39	21	7.0	2458	38	T89399 Methods for diagnosis	5.20e+00
40	21	7.0	2891	10	O57016 PKC delta.	5.20e+00
41	21	7.0	2909	1	O01758 CDNA sequence of rat	5.20e+00
42	21	7.0	8353	38	T89396 Methods for diagnosis	5.20e+00
43	21	7.0	8920	11	O62924 Carbomoyl-phosphate	5.20e+00
44	20	6.7	870	7	O39248 Sequence of cosmid cl	1.45e+01
45	20	6.7	2542	39	V15094 Human apolipoprotein	1.45e+01

ALIGNMENTS

RESULT 1
ID T20017 standard; cDNA to mRNA; 304 BP.
AC T20017.
DT 17-JUL-1996 (first entry)
DE Human gene signature HUMGS01155.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PE 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATSU) MATSUBARA K.
PI (OKUBO) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human CDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing CDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1: Page 538: 22455P; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed CDNA libraries prepared
CC from various human tissues; synthesis of CDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented CDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a CDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.

```

SQ Sequence 304 BP; 84 A; 41 C; 67 G; 106 T;
Query Match 100.0%; Score 298; DB 19; Length 304;
Best Local Similarity 100.0%; Pred. No. 2,92e-173;
Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 gatctaagaagtggtgtttacatggtttacacattcatccttaaaagtgt 60
    |||
OY 1 GATCTAAGCAGGTNGTGTGTACATGTTTACACATTTCCTTAAAGTGT 60

Db 61 tgaagagaagtggttacttaccctcccaaggttgaaagcagggaattcccaagtgctca 120
    |||
OY 61 TGAGAGAGGTTGTATTACCTTCCCAAGGTTGGAAAGCAGGGAATTTCCAGTGTCTTA 120

Db 121 gtttccacagaggaatagtgttaagtagcaagatttgctgcttacctatagtggtc 180
    |||
OY 121 GTTTCCACAGGAGATAGTGTAGTAGCAAAATTTGCTGCTTACATATAGTGTGT 180

Db 181 atgtatgtatatagtaaatngtgtgttaagaagctgatactgatttcatcagcaatg 240
    |||
OY 181 ATGTATGTATATGTAAATNGTGTGTTAAAGAGCTGATCTGATTTTCATAGNCATG 240

Db 241 ttaagcacaagcctccctgcatttgangagcaggtnttcattatattgtattnggga 300
    |||
OY 241 TTAAAGCAAGCCTCCCTGCATTTGANGAGCAGGTNTTCATTATATGATTTTNGGA 300

Db 301 taag 304
OY 301 TAAA 304

RESULT 2
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT primer_bind /function=multiple cloning site
FT 187..204
FT /*tag= b

EP-285123-A.
PD 05-MAY-1988.
PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivu A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prep of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; P: English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 13.4%; Score 40; DB 1; Length 204;
Best Local Similarity 9.6%; Pred. No. 9.57e-10;
Matches 10; Conservative 53; Mismatches 41; Indels 0; Gaps 0;

```

```

Db 87 mrtthyrirmzbnvrydrnsdaaawycvrsykydcacynachdhvvybbyvnhv 146
    |||
CP 289 CATATAATGAANACTGCTCTCTCAATGACAGGAGGCTTGTGCTTACATATGCAATAT 230

Db 147 nhnncncbnnhvcnvhvbnhnmwayrvhndarddhcvcvch 190
    |||
CP 229 GAAATACAGTATACAGCTTCTTAACACACATATTTACATATATATCA 186

RESULT 3
ID Q51746 standard; cDNA; 91 BP.
AC Q51746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
DE Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT ) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI: 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 12.4%; Score 37; DB 9; Length 91;
Best Local Similarity 2.0%; Pred. No. 4.23e-08;
Matches 1; Conservative 43; Mismatches 7; Indels 0; Gaps 0;

Db 10 gssvshyvvhvshvshvshvshvshvshvshvshvshvshvshvshvshvshv 60
    |||
CP 111 GGGAAATTCCTCCCTGCTTCAACCTTGGAAGGTAAATACAACTCTCTCA 61

RESULT 4
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT primer_bind /function=multiple cloning site
FT 187..204
FT /*tag= b

EP-285123-A.
PD 05-MAY-1988.
PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivu A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prep of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; P: English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The

```

```

CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SO Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 11.7%, Score 35; DB 1; Length 204;
Best Local Similarity 7.1%, Pred.No. 5.09e-07;
Matches 7; Conservative 53; Mismatches 39; Indels 0; Gaps 0;

Db 87 mrtthyrmrhmrvdydnrsdaaayccyrrsvkydcydnachddhvybbyvnyh 146
OY 55 AGTTGTTGAGAGAGGTTGATTATACCTTCCCAAGGTGGAACGAGGGAATTTCCCAGT 114
Db 147 nhnncncbhmhcnvbmhnmhnmvaydhdrdvh 185
OY 115 GTCTAGTATTCCACGAGGAAATATGTATAGTACGAA 153

RESULT 5
ID OS1746 standard; cDNA; 91 BP.
AC OS1746:
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT ) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI: 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp: English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (OS1735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also OS1735-45 and OS1747-59.
SO Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 11.1%, Score 33; DB 9; Length 91;
Best Local Similarity 0.0%, Pred.No. 5.91e-06;
Matches 0; Conservative 39; Mismatches 6; Indels 0; Gaps 0;

Db 13 vhsyyvvhvshhsbvhvvhvsvvvvvhvvhvvhvnyh 57
OY 78 ACCTTCCCAAGGTTGGAAGCAGGGGGAATTTCCAGTGCTCTAGT 122

RESULT 6
ID T14322 standard; DNA; 67 BP.
AC T14322:
DT 16-JAN-1997 (first entry)
DE Primer used in the labelling and sorting of nucleotide molecules.
DE Labelling; sorting; sequencing; tag; tagging; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 67
FT /*tag= a
FT /mod_base= Linked to biotin
PN MO9612039-A1.
PD 25-APR-1996.
PF 12-OCT-1995; U12678.
PR 13-OCT-1994; US-322348.
PR 19-DEC-1994; US-359295.

```

(LYNX-) LYNX THERAPEUTICS INC.

Brenner S;
WPI; 96-222023/22.
PT Labelling and sorting moles. using oligo:nucleotide tags - useful in
large-scale parallel operations, e.g. DNA sequencing and mRNA
fingerprinting
PS Disclosure; Page 19; 71pp; English.

Determining the nucleotide sequence (1) of a target polynucleotide
(1) comprises: (a) generating from T a plurality of fragments that
cover T; (b) attaching an oligonucleotide tag from a repertoire of
tags, to each fragment such that all the same fragments have the
same tag, and all different fragments have different tags; (c)
sorting the fragments by specifically hybridising the tags with
their respective tag complements; (d) determining (1) of a portion
of each of the fragments; and (e) determining (1) of T by collating
the sequences of the fragments. The tagging system can be used with
single base sequencing methods to sequence polynucleotides up to
several kilobases in length. The tagging system permits many
thousands of fragments of a target polynucleotide to be sorted onto
one or more solid phase supports and sequenced simultaneously.
This sequence represents a primer which could be used to reverse
transcribe mRNA. If the amplified product is then needed to be
attached to a solid phase support a sequence like the one given in
T14323 may be used. The mRNA would then be removed and the second
strand of cDNA produced using a primer with a similar form to that
described in T14324. After restriction enzyme digestion, the
conjugate would have the formula described in T14325.

Sequence 67 BP; 3 A; 13 C; 4 G; 20 T;

Query Match 8.7%; Score 26; DB 24; Length 67;
Best Local Similarity 26.3%; Pred. No. 2,25e-02;
Matches 10; Conservative 22; Mismatches 6; Indels 0; Gaps 0;

Db 25 www.cwmcwww.cwmcwww.cwmcwww.cwmcwww.cwmcaccgacg 62
:::|||||:|||||:|||||:|||||:|||||

Cp 198 TTACATATATCATCATCATACACTAATATGTAAGCACC 161

RESULT 7
ID T63255 standard; DNA; 68 BP.
AC T63255;
DT 07-OCT-1997 (first entry)
DE Messenger RNA primer containing a complement tag sequence.
KW mRNA; oligonucleotide tag; hybridisation; automated DNA mapping;
OS Synthetic.
FH Key Location/Qualifiers
FT modified_base I
FT /*tag= a
FT /note="5'-labelled with biotin"
FN WO9641011-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09513.
PR 12-OCT-1995; WO-U12791.
PR 07-JUN-1995; US-478238.
PA (SPEC-) SPECTRAGEN INC.
PI Albrecht G, Brenner S;
DR WPI: 97-099943/09.

Sorting poly:nucleotide(s) on to solid supports by attachment to
oligo:nucleotide tags - then specific hybridisation of tags to
immobilised complement, e.g. for automated DNA mapping and
sequencing, genetic identification and diagnosis
PS Disclosure; Page 22; 79pp; English.

A method of sorting a population of polynucleotides on to one or more
solid supports has been produced. The polynucleotides are sorted on to
solid supports by: (a) attaching an oligonucleotide tag to each
polynucleotide, each tag being minimally cross-hybridising; (b) sampling
the population so that all different polynucleotides have different
tags; (c) sorting by specific hybridisation of the tags with their
complements which are attached as uniform populations of identical
oligonucleotides in spatially distinct regions on one or more solid
phase support. The present sequence represents a primer, from a set,
preferably containing complements of tag sequences. When the population

Db **a** accacwmcwwwcwwcwwwcwwwcwwwcwwwc wwwc w
Cp 203 ACNATTACATATATACATACATACACACTTAT 172

RESULT 8
ID T73397 standard; DNA; 68 BP.
AC T73397;
DT 03-DEC-1997 (first entry)
DE Oligonucleotide tag containing primer #1 for toxicity determination.
KW PCR; primer: amplify polymerase chain reaction; toxicity determination;
OS oligonucleotide tag; hepatocyte mRNA populations fingerprinting; ss.
FH Key
FT modified_base Location/Qualifiers
TT 1
TT /*tag= a
TT /note= "biotin labelled"
TT misc_feature 12..47
TT /tag= b
TT /note= "Oligonucleotide tag"
PN MO9713877-A1.
PD 17-APR-1997.
PF 11-OCT-1996; U16342.
PR 06-JUN-1996; WO-009513.
PR 12-OCT-1995; WO-012791.
PA (LYNX-) LYNX THERAPEUTICS INC.
PI Martin DW:
DR WPI: 97-235911/21.
PT Massively parallel signature sequencing - useful to test toxicity of compound, or to identify genes which are differentially expressed in selected tissue or a test animal after treatment with a compound
PS Disclosure; Page 26; 65pp; English.
CC T73397-T73400 represent amplification primers that can be used in the method of the invention. The method is for determining the toxicity of a compound. The method comprises administering a compound to a test organism, extracting RNA molecules from 1 or more tissues, and forming a population of cDNA molecules from each RNA population. Each cDNA molecule has attached an oligonucleotide tag, and then each population of cDNA molecules is then sampled. The cDNA molecules are then sorted by specifically hybridising the tags with their complements, which are attached in spatially discrete regions on 1 or more solid phase supports. CC The sequence of each of the sorted cDNA molecules is determined to form a frequency distribution of expressed genes for each tissue. The frequency of distribution of expressed genes is then correlated with the toxicity of the compound. The method, which comprises the massively parallel signature sequencing technique, can be used to test the toxicity of a compound, or to identify genes which are differentially expressed in a selected tissue of a test animal after treatment with a compound, in a mammalian tissue culture, preferably comprising hepatocytes. The method may also be used to fingerprint mRNA populations, either in isolated measurements or in the context of dynamically changing population,
CC partial sequence information is obtained from a large sample, e.g. 10 to 100000, or more, of cDNA attached to separate microparticles.

```

Query Match          8.4%; Score 25; DB 34; Length 68;
Best Local Similarity 25.0%; Pred. No. 6,93e-02;
Matches      8; Conservative    20; Mismatches   4; Indels    0; Gaps    0;

Db      8 accacgwwcwwwcwwwcwwwcwwwcwwwcwww 39
CP      203 ACNATTACATATATACATACATACACTACT 172
        |||:::||||:||||:||||:||||:||||:
RESULT  9
ID      T14325 standard; DNA; 70 BP.
AC      T14325:
DE      16-JAN-1997 (first entry)
KW      Conjugate formed by labelling and sorting of nucleotide monomers.
OS      Labelling; sorting; sequencing; tag; tagging; ss.
FH      Key
FT      misc_feature
FT      Location/Qualifiers
FT      /tag= a
FT      /note= "This N represents the amplified cDNA
           sequence"
PN      WO9612039-A1.
PD      25-APR-1996.
PF      12-OCT-1995; U12678.
PR      13-OCT-1994; US-322348.
PR      19-DEC-1994; US-359295.
PA      (LTX-) LTX THERAPEUTICS INC.
PI      Brenner S;
DR      WPI: 96-222023/22.
PT      Labelling and sorting mols. using oligo:nucleotide tags - useful in
PT      large-scale parallel operations, e.g. DNA sequencing and mRNA
PT      fingerprinting
PS      Disclosure; Page 20; 71pp; English.
PP      Determining the nucleotide sequence (1) of a target polynucleotide
CC      (T) comprises: (a) generating from T a plurality of fragments that
CC      cover T; (b) attaching an oligonucleotide tag from a repertoire of
CC      tags, to each fragment such that all the same fragments have the
CC      same tag, and all different fragments have different tags; (c)
CC      sorting the fragments by specifically hybridising the tags with
CC      their respective tag complements; (d) determining (1) of a portion
CC      of each of the fragments; and (e) determining (1) of T by collating
CC      the sequences of the fragments. The tagging system can be used with
CC      single base sequencing methods to sequence polynucleotides up to
CC      several kilobases in length. The tagging system permits many
CC      thousands of fragments of a target polynucleotide to be sorted onto
CC      one or more solid phase supports and sequenced simultaneously.
CC      A primer which is initially used to reverse transcribe mRNA is
CC      described in T14322. If the amplified product is then needed to be
CC      attached to a solid phase support a sequence like the one given in
CC      T14323 may be used. The mRNA would then be removed and the second
CC      strand of cDNA produced using a primer with a similar form to that
CC      described in T14324. After restriction enzyme digestion, the
CC      conjugate would have a formula similar to this.
SQ      Sequence       70 BP;      2 A;      12 C;      3 G;      19 T;

Query Match          8.4%; Score 25; DB 24; Length 70;
Best Local Similarity 25.0%; Pred. No. 6,93e-02;
Matches      8; Conservative    20; Mismatches   4; Indels    0; Gaps    0;

Db      4 accacgwwcwwwcwwwcwwwcwwwcwwwcwww 35
CP      203 ACNATTACATATATACATACATACACTACT 172
        |||:::||||:||||:||||:||||:||||:
RESULT  10
ID      T172882 standard; cDNA; 19124 BP.
AC      T172882:
DE      12-SEP-1997 (first entry)
DE      Plasmodium var-7 gene.
KW      DBL gene family; SAMP; sialic acid binding protein; vaccine; therapy;
KW      Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
KW      DABP; mezozyote; malaria; var-1; var-2; var-3; var-7; immune response;

```


[illegible]

Dc domain with affinity for a ligand and a second effector peptide portion that is chemically or biologically active. They may further comprise a linker peptide between the 2 domains. The oligonucleotides are also designed so that the expressed peptide contains 2 or 4 cysteine residues positioned in, or flanking, the unpredicted or variant residues. These residues confer some degree of conformational rigidity to the peptides.

The TSARs or comps. comprising a TSAR binding domain can be used in vivo to deliver a chemically or biologically active moiety, eg., metal ion, radiotope, peptide, toxin or enzyme, to the specific target or on the cell. They can also replace the function of macromolecules, eg., monoclonal or polyclonal antibodies and therefore circumvent the need for complex methods of hybridoma formation or in vivo antibody production. The TSARs are easily characterised and have been designed actively allowing direct and rapid detection in a screening process.

SQ Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;

Query Match Best Local Similarity 7.7%; Score 23; DB 12; Length 114; Matches 8; Conservative 23; Mismatches 75; Indels 0; Gaps 0

Dc 6 bmbnbnbnbnbnbnbnbnbnbnbtgccttcgcnbnbnbnbnbnbnbnnnnnbnbn 65
|.....|
Dy 129 CCAGGGAATGTCTTAACGTCGAAAGTATTGGCTCCACTAATGTTGATACT 188
|.....|
Db 66 bnbnbnbnbnbnbnbnbnbnbnbtgcnbbnbnbnbnbnbnbnbnbnbnbnbn 111
|.....|
Dy 189 ATATATGTAAATNGTGCTTAAGAAGCGTAGACTGATTTGCATATG 234
|.....|

RESULT 14
ID T20017 standard; cDNA to mRNA; 304 BP.
AC T20017;
DT 17-JUL-1996 (first entry)
DE Human gene signature HMGSO1155.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency; human; cloning; mapping; non-biased library; diagnosis; detection; KJ cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATSU) MATSUBARA K.
PI (OKUBO) OKUBO K.
PM Matsubara K.; Okubo K.; WPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g. for diagnosis of abnormal cell function, by preparing cDNA that reflects relative abundance of corresp. mRNA in specific human tissues
PS Claim 1: Page 538; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp. double-stranded DNA) which comprises one of the 7837 "GS" sequences given in TI9001-T26837 and which is able to hybridize to part of human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature) sequences were obtained from 3'-directed cDNA libraries prepared from various human tissues; synthesis of cDNA was initiated from the 3' end of mRNA by using poly(T) as the sole primer. Since the 3'-untranslated sequence is unique to a particular mRNA species, almost all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library is constructed so as to reflect accurately the relative abundance of different mRNAs in the particular tissue from which it was derived. The appearance frequency of a given GS in a cDNA library can be determined (esp. using primers and probes derived from the GS sequences) as a means of diagnosing abnormal cell function or for recognising different cell types.
SQ Sequence 304 BP; 84 A; 41 C; 67 G; 106 T;

Query Match Best Local Similarity 7.7%; Score 23; DB 19; Length 304; Matches 30; Conservative 7; Mismatches 7; Indels 0; Gaps 0

Dc 138 tatgtataatgccaaattacttcttaaccata 174

CP 174 TATATGTAAGACCAATACCTTCTACTACACATA 138

RESULT 15

ID T43928 standard: DNA; 1380 BP.

AC T43928:

DT 18-AUG-1997 (first entry)

DE Sequence around marker 184 in HH region of chromosome 6p2.1.

KW Primer: polymerase chain reaction; amplify; hereditary haemochromatosis;

HH: mutation; HH-associated allele; base-pair polymorphism; HHP-1;

KW HHP-19; HHP-29; microsatellite repeat allele; genetic marker;

KW interferon treatment; hepatitis C infection; ss.

OS Synthetic.

FH key Location/Qualifiers

FT repeat_region 641..672

FT /tag= a

FT /note= "Dinucleotide repeat region"

PN MO9635803-A1.

PD 14-NOV-1996

PF 08-MAY-1996: U06583.

PR 08-MAY-1995; US-436074.

PR 15-NOV-1995; US-559302.

PR 09-FEB-1996; US-599252.

PA (MERC-) MERCATOR GENETICS INC.

PI Drayna DT, Feder JN, Gailke A, Kimmel BF, Thomas WJ;

PI Wolf RK;

PI WPI; 96-518691/51.

PT Diagnosing and genotyping of hereditary haemochromatosis (HH) -

PT using primers to detect specific polymorphisms of the HH gene on

PT chromosome 6p2.1 or novel microsatellite markers

PS Claim 24: Fig 1D: 67pp: English

CC The sequences given in T43925-55 represent portions of the genome

CC surrounding several markers of the invention. The markers were

CC identified using the series of primer pairs given in T71901-72

CC which were used to determine the presence or absence of the common

CC hereditary haemochromatosis (HH) gene mutation in an individual. The

CC method comprised assessing genomic DNA from an individual for the

CC presence or absence of the HH-associated allele of the single base-pair

CC polymorphism HHP-1, HHP-19 or HHP-29, and/or at least one non-optional

CC marker comprising the following microsatellite repeat alleles of group

CC A and optionally of group B:

CC Group A: 19P9, 18B4, 1A2, 1E4, 24E2, 2B8, 3321-1, 4073-1, 4440-1, 4440-2,

CC 731-1, 5091-1, 3216-1, 4072-2, 950-1, 950-2, 950-3, 950-4, 950-5, 950-6,

CC 950-8, 63-1, 63-2, 63-3, 65-1, 65-2, 373-8, 373-29, 68-1, 241-6, 241-29;

CC Group B: D6S464, D6S306, D6S258, D6S265, D6S105 and D6S1001.

CC The absence of the genotype indicates the likelihood of the presence of

CC the HH mutation. Knowledge of the new genetic markers allows the

CC definition of genotypes characteristic of heterozygous carriers and

CC homozygotes having a HH mutation in their genomic DNA. The potential for

CC HH in an individual interferes with the effectiveness of interferon

CC treatment for hepatitis C infection. By diagnosing this potential, the

CC responsiveness of interferon treatment may be evaluated.

SO Sequence 1380 BP: 364 A: 261 C: 186 G: 569 T:

Query Match 7.7%: Score 23; DB 30; Length 1380;

Best Local Similarity 81.1%: Pred. No. 6.24e-01;

Matches 30; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 615 gcttacaatagagtggtgtgtatataatat 651

Qy 164 GCTTACATATAGTGTATGATATATATATAT 200

Search completed: Tue Dec 22 04:36:31 1998
Job time : 68 secs.


```

exon      872..1000
          /gene="Cp1"
          /number=1
intron    1001..2309
          /gene="Cp1"
          /number=2
exon      2310..2426
          /gene="Cp1"
          /number=2
CDS       join(2328..2426,6476..6690,6751..7462)
          /gene="Cp1"
          /codon_start=1
          /product="cysteine proteinase-1"
          /db_xref="pid:g2305221"
          /translation="MRTAVLPLALIAVAQVSPADYVMEEMHFKLEHKNYDDE
BERFLKTFENKKHIAKHNOFPARGSKYFKLAIVKYLADLHEEROLMNGFNYLHK
OLRADESEFKVETLSPAHVTLPKSVDMRTKGAVAVRDQCGSCMAFSSGALGEO
HFRKSGVLVLSLEQMLVDCSTKYGNNGCNGGLMDNAFRYIKDNGCIDKESYPAID
DSCHFNKGATDGRFTDIPQDDEKMAEAVATGPVSAIDAHSSEFOFSEGVN
EPQCDANLDHGVLVVGFTDSEGDYLVKNKSWGTGDKGFIKMLNKENGCIAS
ASSYPLV"
intron    2427..6475
          /gene="Cp1"
          /number=3
misc_feature 4546..4553
          /gene="Cp1"
          /note="insertion site of P(Casper)(50c)"
          /citation=[1]
exon      6476..6690
          /gene="Cp1"
          /number=3
intron    6691..6750
          /gene="Cp1"
exon      6751..7707
          /gene="Cp1"
          /number=4
mRNA      join(<8110..9300,9370..>9532)
          /product="phenylalanyl tRNA synthetase"
          /join(8110..9300,9370..>9532)
          /note="potential orf"
          /codon_start=1
          /product="phenylalanyl tRNA synthetase"
          /db_xref="pid:g2305222"
          /translation="MLTLRVOGARHMLKSTRCLASAPAKSPSPPOLEYSSTVA
TDGWTNVPKILSYGANKHLODHPLSIIRIARYFVGARNGNDLPSFYDGM
PVYVQONDNLIPADHYRSROKSDCYIINOCHILRAHTAHOVELIGGIDNPELVG
EYVRDEIDSTHYFVPHQADAVRLTKDKLEFRNNGELFEETNSGTLADPKLILPH
SSWTPNSPAPTRAVKLEHEKHHVLCVLTDLFGPRIKRYVDVTFEFTDPSWLEIT
YFKDWLLEVLGGIMRHETLORSGVHOSIGVAFGLERLAVLFDIDIRLFWSNDS
GFLDSFSEKDLNLPKRPISHYPOCTNDLSFWLPODIEVDAGFSPNDPYDIYRVSAG
DMVEQISLVDFKHKPKTKGSSVCFRIYVRHMERLTQAQEVNEIHRQIASASVDSFNVQ
IX"
BASE COUNT 2929 a 2357 c 2282 g 3046 t 158 others
ORIGIN
Query Match 12.4%; Score 37; DB 18; Length 10772;
Best Local Similarity 14.2%; Pred. No. 2,32e-06;
Matches 17; Conservative 63; Mismatches 39; Indels 1; Gaps 1;
Db 1704 WKMMWTTWKMAAKYRYTMMKMYTSRTTTSAMMMWYTWSTWRYWMAVAMKMMWT 1763
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cp 237 TGNCAATATGAATAATGATGCTTTAACACACNATTTAC-ATAATATACATACATAC 179
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1764 RTMARANAMASNRMKWTASAAAYASARMKMMKMYARAKKTTMMAAMAKMKAAAMWTR 1823
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cp 178 ACACTATATGTAAGCAGCAAAATACCTTGCTACTTACACAAATTCCTGCTGTGGAACAATA 119
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS     Primates; Catarrhini; Hominoidea; Homo.
            1 (bases 1 to 216021)
            Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavag,A.R. and Venter,J.C.
            Homo sapiens Chromosome 16 BAC clone C1198/5K-A-952F10
            Unpublished
            2 (bases 1 to 216021)
            Adams,M.D. and Loftus,B.J.
            Direct Submission
            Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, Email:
            bjoftus@tigr.org
            3 (bases 1 to 216021)
            Adams,M.D. and Loftus,B.J.
            Direct Submission
            Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT      On Jul 24, 1998 this sequence version replaced gi:3241936.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org. The orientation of the sequence is from
            Sp6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), GeneIndex (Phil Green, University of
            Washington), Genscan (Chris Burge,
            http://genomic.stanford.edu/~chris/GENSCANW.html) searches of the
            complete sequence against a peptide database, and the Human gene
            index database at tigr (http://www.tigr.org/tdb/ngi/ngi.html).
            Genes without peptide homology having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES     source
            1..216021
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="16"
            /map="16q21-22"
            /clone="A-952F10"
            27765..27872
            /note="7766, STS1-csRL-27g3-uA/csRL-27g3-uZ, Chr. -, Homo
            sapiens"
            /db_xref="dbSTS:G02122"
            175801..175945
            /note="16084, CHIC.GCT10B02, Chr. -, Homo sapiens"
            /db_xref="dbSTS:G09703"
            175810..175945
            /note="16316, CHIC.GCT15C04, Chr. -, Homo sapiens"
            /db_xref="dbSTS:G09935"
            199463..199572
            /note="9824, WI-3555, Chr. 16, Homo sapiens"
            /db_xref="dbSTS:G04338"
BASE COUNT 60960 a 51778 c 49172 g 53987 t 124 others
ORIGIN
Query Match 12.1%; Score 36; DB 26; Length 216021;
Best Local Similarity 14.1%; Pred. No. 8,20e-06;
Matches 19; Conservative 66; Mismatches 50; Indels 0; Gaps 0;
Db 1442 TYCCYYCSRRRRGKKKMYMSRRARRRSGAKKKKYYYYCYYYYYCYCWRAMMA 1501
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cp 137 TTCCTCTGTGTGGAACACTGACACTGGAAATTCCTCCTTCCAACTTGGGAAGCT 78
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1502 AAMYKRRSCMAYYMYNMGRAKTYYYMARRRGCARSYKMYAMWTWMAAMTCAAAAANA 1561
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cp 77 AATACAACTCTCTCAACAACCTTTTAAAGATGAATGTGTGTAACAACTGTAAACAC 18
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Db 1562 MYTTYMMMYKKW 1576
 CP 17 ACNACTGCTTAGA 3

RESULT 3 AF012089 10772 bp DNA INV 05-AUG-1997
 LOCUS Drosophila melanogaster cysteine proteinase-1 (Cp1) gene, complete
 DEFINITION cds, and phenylalanyl tRNA synthetase gene, partial cds.
 ACCESSION AF012089
 MID 92305220
 KEYWORDS fruit fly.
 SOURCE Drosophila melanogaster
 ORGANISM Eukaryota; Mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachyera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 4546 to 4553)
 AUTHORS Gray, Y.H., Tanaka, M.M. and Svred, J.A.
 TITLE P-element-induced recombination in Drosophila melanogaster: hybrid element insertion
 JOURNAL Genetics 144 (4), 1601-1610 (1996)
 MEDLINE 97132596
 REFERENCE 2 (bases 1 to 10772)
 AUTHORS Gray, Y.H.M., Svred, J.A., Preston, C.R. and Engels, W.R.
 TITLE Structure of the cysteine proteinase (Cp1) gene of Drosophila melanogaster and associated mutational effects
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 10772)
 AUTHORS Gray, Y.H.M., Svred, J.A., Preston, C.R. and Engels, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (30-JUN-1997) School of Biological Sciences, University of Sydney, Biology A12, Sydney University, NSW 2006, Australia
 FEATURES
 source 1..10772
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 join(872..1000,2310..2426,6476..6690,6751..7707)
 /gene="Cp1"
 /product="cysteine protease"
 gene 872..7707
 /gene="Cp1"
 exon 872..1000
 /gene="Cp1"
 intron /number=1
 1001..2309
 /gene="Cp1"
 exon 2310..2426
 /gene="Cp1"
 CDS /number=2
 join(2328..2426,6476..6690,6751..7462)
 /gene="Cp1"
 /codon_start=1
 /product="cysteine proteinase-1"
 /db_xref="pid:92305221"
 /translation="MRTAVLLPLALLAAVAQVSVADVMEEMHTFKLEHRKNYODET
 EERFLKIFENKRIKIAHNOFAEGKYSFKLAVKADLLHHEFRQLANGNYLHK
 QLRADSEFKGVTFISPAHVLPKSVDMRTGAVAVADOGCGSCAFSSGALGEG
 HFRKSGVLVSEONLVDCSTRYGNNGGIMADAFRYIKDNGSIDKESVPEYAI
 DSCHEKGTGATDGTGFDIPOGDKKMAEAVATYGPVSALIDASHESFOYSEVYN
 EPQCDKAOHLDHGCVLVYVGFTDSEGDYLVKNSMGTGDKGFTMLNKKENOCGIAS
 ASSTPLV"
 intron 2427..6475
 /gene="Cp1"
 misc_feature 4546..4553
 /gene="Cp1"
 /note="insertion site of P(casper)(50C)"
 exon 6476..6690
 /gene="Cp1"
 /number=3
 intron 6691..6750
 /gene="Cp1"

exon 6751..7707
 /gene="Cp1"
 mRNA /number=4
 join(<8110..9300,9370..>9532)
 /product="phenylalanyl tRNA synthetase"
 CDS join(8110..9300,9370..>9532)
 /note="potential orf"
 /codon_start=1
 /product="phenylalanyl tRNA synthetase"
 /db_xref="pid:92305222"
 /translation="MLTLRYOGARHMLKSRCLASSAAPAKSPSPPOLEVGSTYA
 TDGWTNTPKRLISYVGANKHIOHDPISLIIRORIVANYEYGRNORGNPLESVYDKN
 PVYVQONFENLLIPADHVSRSKSDCYINQHLRHATTHAQVELISGCDLNLVVG
 EYVRDEIDSTHYFVHQADKRVLTLDKLEERNPGLLEETWSGLDKLILPH
 SSWTKPNSPARRAVKLMEHEMKHVLGTLKDLGPRIKRYMVDYFEPFQPSWELEI
 YFKNMLEVLGCGIMRHEILORSGVHOSIGYAFGLERLAMYFDIPILEFNSDS
 GFLQSEKDLHNP.KYKPISHYPOGINDSEFMLODIEVAGVSPNDEYLVRSVAG
 DMVFOISLVKDFKPKKSGVCRIRYRHHERTLTQAEVNEIHKQIASASVDSFNVO
 IX"

BASE COUNT 2929 a 2357 c 2282 g 3046 t 158 others
 ORIGIN

Query Match 11.1%; Score 33; DB 18; Length 10772;
 Best Local Similarity 17.3%; Pred. No. 3.35e-04;
 Matches 24; Conservative 64; Mismatches 50; Indels 1; Gaps 1;

Db 1694 WYVWMTTMMKMMWTKKMAKRYRTWMMKMYTSTRTTTSAMMMYTWSTWTKYMW 1753
 QY 74 ATTACCTTCCCAAGTGTTGAAGAAGGGAATTCCAGTCTCTAGTTTCCAC-CAG 132
 Db 1754 AYAMKMMWTRTWMANANSAARWKTSAAAYAMKMMKWYRAMKTKTMMAMKW 1813
 QY 133 AGCAATATGCTGAACAGCAAGTATTGCTCTACATATGTTGATGATGTATAT 192

Db 1814 RWKAAMTWTRWMTTAA 1832
 QY 193 ATGTAATGTTGTATAA 211

RESULT 4
 LOCUS HSIQKLO9 2926 bp DNA PRI 25-JUL-1994
 DEFINITION H.sapiens gene for immunoglobulin kappa light chain variable region
 O9 and O10.
 ACCESSION X71896
 MID 9434975
 KEYWORDS immunoglobulin; immunoglobulin kappa light chain variable region.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2926)
 AUTHORS Schaeble, K.F.
 TITLE Direct Submission
 JOURNAL Submitted (04-MAY-1993) K.F. Schaeble, Institut f Physiologische Chemie, der Universitaet Muenchen, Schillerstr 44, 8000 Muenchen, FRG

FEATURES
 source 1..2926
 location/Qualifiers
 /organism="Homo sapiens"
 /isolate="placenta 5c"
 /db_xref="taxon:9606"
 /tissue_type="placenta"
 /clone_lib="cosmid"
 /map="p12"
 /chromosome="2"
 gene 465..1730
 /gene="O9"
 misc_feature 465
 /gene="O9"
 exon 578..632
 /note="start of de box"
 /gene="O9"

```

intron      /number=1
            633..750
            /gene="09"
exon        /number=1
            751..1052
            /gene="09"
            /number=2
            1055
misc_feature /gene="09"
            /note="start of hepta box"
            1064
misc_feature /gene="09"
            /note="start of nona box"
            1505..1730
repeat_region /gene="09"
            /note="alu repetitive sequence"
            1755..2416
            /gene="010"
            1755
misc_feature /gene="010"
            /note="start of dc box"
            1857..1905
exon        /gene="010"
            /number=1
            1906..2100
            /gene="010"
intron      /number=1
            2101..2394
            /gene="010"
            /number=2
            2397
misc_feature /gene="010"
            /note="start of hepta box"
            2416
misc_feature /gene="010"
            /note="start of nona box"
            /note="start of nona box"
            842 a 649 c 632 g 803 t
BASE COUNT 842 a 649 c 632 g 803 t
ORIGIN
Query Match 9.7%; Score 29; DB 25; Length 2926;
Best Local Similarity 86.8%; Pred. No. 3.87e-02;
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 2724 ACACACTATATATATATATACACACACTATAT 2761
Cp 207 ACACACATTTACATATACATACACACTATAT 170

RESULT 5
LOCUS HS733D15 40577 bp DNA PRI 02-OCT-1998
DEFINITION Human DNA sequence from clone 733D15 on chromosome Xp11.3. Contains
a Zinc-finger (pseudo?) gene and GSSs, complete sequence.
ACCESSION AL031393
NID g3646112
KEYWORDS HTG: Zinc-finger.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 40577)
BIRD.C.
REFERENCE Submitted (24-SEP-1998) Sanger Centre, Hinxton, Cambridgeshire,
AUTHORS CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk
TITLE Requests: clonerequests@sanger.ac.uk
JOURNAL On Sep 24, 1998 this sequence version replaced g1:3550752.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
COMMENT IMPORTANT: This sequence is not the entire insert of clone 733D15.

```

It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The true left end of clone 711L6 (AL022165) is at 40478 in this sequence.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>

733D15 is from the library RPC14 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:PCRYAC2>.

FEATURES

```

SOURCE
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="733D15"
/map="p11.3"
/clone_lib="RPC14"
complement(1..58)
/note="Alub repeat: matches 62..5 of consensus"
154..245
/note="Alub repeat: matches 1..92 of consensus"
278..632
/note="THE1C repeat: matches 31..371 of consensus"
633..782
/note="THE1B-INTERNAL repeat: matches 1..156 of consensus"
780..970
/note="THE1B-INTERNAL repeat: matches 97..290 of consensus"
972..1277
/note="AlusX repeat: matches 1..302 of consensus"
1279..1368
/note="THE1B-INTERNAL repeat: matches 286..375 of consensus"
1369..1416
/note="24 copies 2 mer tt 85% conserved"
complement(1427..2314)
/note="LIP2 repeat: matches 889..1 of consensus"
complement(2165..3344)
/note="L1 repeat: matches 5390..4158 of consensus"
3348..3855
/note="MER9 repeat: matches 3..510 of consensus"
complement(3844..6211)
/note="L1 repeat: matches 4175..1827 of consensus"
6212..6505
/note="AluY repeat: matches 1..294 of consensus"
complement(6506..8335)
/note="L1 repeat: matches 1832..14 of consensus"
complement(8338..11334)
/note="remnants of endogenous retrovirus"
complement(11509..12113)
/note="MER1B repeat: matches 632..16 of consensus"
12399..12628
/note="MTR1E repeat: matches 264..515 of consensus"
12629..12921
/note="AlusX repeat: matches 1..294 of consensus"
12925..12986
/note="MTR1E repeat: matches 508..568 of consensus"
14044..14228
/note="L1 repeat: matches 5201..5390 of consensus"
14085..14924
/note="LIP10 repeat: matches 1..911 of consensus"
15413..15547
/note="AluIO repeat: matches 1..136 of consensus"
15548..15842
repeat_region

```



```

repeat_region /note="AluJo repeat: matches 2. .298 of consensus"
15845. .16023
/note="L1 repeat: matches 3881. .4054 of consensus"
repeat_region complement(16055. .16558)
/note="AluJo repeat: matches 290. .87 of consensus"
repeat_region 16269. .16312
/note="22 copies 2 mer 19 96% conserved"
repeat_region complement(16321. .16451)
/note="AluB repeat: matches 129. .1 of consensus"
16944. .17114
/note="MLT2-internal repeat: matches 2249. .2432 of
consensus"
repeat_region 17134. .17360
/note="MLT2-internal repeat: matches 2938. .3168 of
consensus"
repeat_region 17382. .17458
/note="MLT2-internal repeat: matches 3224. .3299 of
consensus"
repeat_region 17471. .17673
/note="MLT2-internal repeat: matches 2994. .3199 of
consensus"
repeat_region complement(17711. .18080)
/note="THE1B repeat: matches 364. .1 of consensus"
repeat_region complement(18525. .18824)
/note="AluJo repeat: matches 302. .1 of consensus"
repeat_region complement(19145. .19297)
/note="THE1B repeat: matches 364. .210 of consensus"
repeat_region complement(19296. .19388)
/note="MER28 repeat: matches 93. .2 of consensus"
20431. .20737
/note="AluSg repeat: matches 3. .298 of consensus"
20832. .20955
/note="FLAM_C repeat: matches 7. .130 of consensus"
repeat_region complement(20990. .21292)
/note="AluJo repeat: matches 301. .2 of consensus"
gene complement(21669. .23161)
/gene="dJ733D15.1"
complement(21669. .23161)
/gene="dJ733D15.1"
/note="could be (part of) a pseudogene; match: proteins
015918 P51523 Q14587 Q06730 P15620 Q14585 P51522 P51508
P17141 Q62510 P18732 Q14593 Q43693 Q43724 Q43340
Q03923 Q14709 P18744 Q62788 Q43345 Q60893 P51814 Q26618
P08045 Q43338 P51505 P51507 P38621; match: cDNA D31763"
/codon_start=3
/evidence=not_experimental
/product="dJ733D15.1 (Zinc-finger protein)"
/db_xref="PID:el329909"
/db_xref="PID:g3702137"
/translation="VWEVDEQIDHYKESQDKLQQAATIGKTTLKDESGQCKICKRI
IYLNDFVSVKQRLPKRYISWERSNHNFLQNSYARKKDGKAYWKVCLAHYLNH
KAOPAEERFDNKGALHOKQALRSQSGEGLKQTEGKAFIOKANLVHORHT
HTGEPEYCECEAKAFSOKSTLIHQHRTGKPYECSCGTFLOKSTLIHQHRT
GKPEVDCGCPKAFKSYHLIRHEKTHIRQAFKGIKCTSLIYORHTSPCCSE
HGKASDESPKTHRTKENTYKESCGKSPKSHSVQRHTHTEKPEYCEICG
KTPSKSHSVHRTHTGKPEYECRCGKAFGEKSTLIHQHMGHGEKPYKONEGKA
FSEKSLIKHQRHTGKPEYECTDCKKAFSRSTLIHQHRTGKPYKCSBGKAFS
VKSTLIHRTHTGKPEYECRCGKAFSGKSTLIHQHSHGDKML"
22234. .22550
/misc_feature /gene="dJ733D15.1"
/note="match: GSS B17284"
23852. .24151
repeat_region /note="AluSg repeat: matches 1. .300 of consensus"
25667. .25874
repeat_region /note="AluB repeat: matches 86. .292 of consensus"
complement(26172. .26468)
/note="AluSg repeat: matches 297. .1 of consensus"
repeat_region complement(26714. .27013)
/note="AluSg repeat: matches 300. .1 of consensus"
27316. .27598
repeat_region /note="AluSg repeat: matches 2. .284 of consensus"
27599. .27620
/note="11 copies 2 mer ca 100% conserved"

```

..
Note: remainder of annotations omitted.

Query Match 9.4%; Score 28; DB 27; Length 40577;

Best Local Similarity 73.3%; Pred. No. 1.22e-01;

Matches 44; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```

Db 35853 TGCTTCACGCCCTGACACAGAGAGACCTCTCTCAAAAATTTTTAAAGAAAGAA 35912
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
Cp 99 TGCTTTCACACTTGGGAGGTAAATCAACACTCTCAACAACTTTTAAAGATGAA 40

```

```

RESULT 6
LOCUS HUM1GKVS 895 bp DNA PRI 05-JAN-1995
DEFINITION Human germline pseudo-Igk chain gene, V301-region.
ACCESSION M17765
NID 9186034
KEYWORDS C-region; V-region; germline; immunoglobulin light chain;
immunoglobulin-kappa; pseudogene; variable region subgroup VK3.
Homo sapiens foetus liver DNA.
SOURCE
ORGANISM Homo sapiens
Eukaryote; mitochondria; eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 895)
REFERENCE
AUTHORS Chen,P.P., Albrandt,K., Kipps,T.J., Radoux,V., Liu,F.T. and
Carson,D.A.
Isolation and characterization of human VKIII germ-line genes.
Implications for the molecular basis of human VKIII light chain
diversity
JOURNAL J. Immunol. 139 (5), 1727-1733 (1987)
MEDLINE 87309783
COMMENT Computer-readable copy of sequence in [1] kindly provided by
P.Chen, 08-DEC-87.
FEATURES
source Location/Qualifiers
1..895
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="foetus"
/germline
/tissue_type="liver"
/map="22q11"
85..379
/gene="IGKVP3"

```

```

CDS
    <85..>379
    /gene="IGKVP3"
    /note="V-region"
    /pseudo
    /db_xref="GDB:G00-120-091"
    /codon_start=1
    382..388
    /note="heptamer recombination signal"
    /organism="Homo sapiens"
    401..409
    /note="nonamer recombination signal"
    /organism="Homo sapiens"
    misc_recomb
        166 c 178 g 268 t
BASE COUNT      283 a
ORIGIN
    9.1%; Score 27; DB 25; Length 895;
    Best Local Similarity 84.2%; Pred. No. 3.77e-01;
    Matches 32; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 693 ACGCACTATATATATATATACACACACTATAT 730
    ||||| ||||| ||||| ||||| ||||| |||||
Cp 207 ACACACNMTTTCATATATACATACACACTATAT 170

RESULT 7
LOCUS HUMIGVR 905 bp DNA PRI 05-JAN-1995
DEFINITION Human germline pseudo-Igk chain gene, VJ10-region.
ACCESSION M17764
NID g166033
KEYWORDS C-region; V-region; germline; immunoglobulin light chain;
    immunoglobulin-kappa; pseudogene; variable region subgroup VK3.
SOURCE Homo sapiens
    Homo sapiens foetus liver DNA.
ORGANISM Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
    Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 905)
AUTHORS Chen,P.P., Albrandt,K., Kipps,T.J., Radoux,V., Liu,F.T. and
    Carson,D.A.
TITLE Isolation and characterization of human VKII germ-line genes.
    Implications for the molecular basis of human VKII light chain
    diversity
JOURNAL J. Immunol. 139 (5), 1727-1733 (1987)
MEDLINE 87309783
COMMENT Computer-readable copy of sequence in [1] kindly provided by
    P.Chen, 08-DEC-87.
FEATURES
    source
        1..905
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /dev_stage="foetus"
        /germline
        /tissue_type="liver"
        /map="22q11"
        85..379
        /gene="IGKVP3"
        <85..>379
        /gene="IGKVP3"
        /note="V-region"
        /pseudo
        /db_xref="GDB:G00-120-091"
        /codon_start=1
        382..388
        /note="heptamer recombination signal"
        /organism="Homo sapiens"
        401..409
        /note="nonamer recombination signal"
        /organism="Homo sapiens"
    misc_recomb
        165 c 176 g 273 t
BASE COUNT      291 a
ORIGIN
    9.1%; Score 27; DB 25; Length 905;
    Best Local Similarity 84.2%; Pred. No. 3.77e-01;
    Matches 32; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Db 703 ACACATATATATATATATATACACACAGTATAT 740
    ||||| ||||| ||||| ||||| ||||| |||||
Cp 207 ACACACNMTTTCATATATACATACACTATATAT 170

RESULT 8
LOCUS HSIKTLA15 3083 bp DNA PRI 25-JUL-1994
DEFINITION H.sapiens gene for immunoglobulin kappa light chain variable region
    A15 and A16.
ACCESSION X71883
NID g434959
KEYWORDS immunoglobulin; immunoglobulin kappa light chain variable region.
SOURCE human.
ORGANISM Homo sapiens
    Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
    Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3062)
AUTHORS Schaeble,K.F.
TITLE Direct Submission
JOURNAL Submitted (04-MAY-1993) K.F. Schaeble, Institut f Physiologische
    Chemie, der Universitaet Muenchen, Schillerstr 44, 8000 Muenchen,
    FRG
REFERENCE 2 (bases 1 to 3083)
AUTHORS Schaeble,K., Thiebe,R., Flugel,A., Weindl,A. and Zachau,H.G.
TITLE The human immunoglobulin kappa locus: pseudogenes, unique and
    repetitive sequences
JOURNAL Biol. Chem. Hoppe-Seyler 375 (3), 189-199 (1994)
MEDLINE 94280631
FEATURES
    source
        Location/Qualifiers
        1..3083
        /organism="Homo sapiens"
        /isolate="Placenta St"
        /db_xref="taxon:9606"
        /tissue_type="placenta"
        /clone_lib="cosmid"
        /clone="cos237"
        /map="p12"
        /chromosome="2"
        426
        /gene="A15"
        /note="start of dc box"
        426..1691
        /gene="A15"
        539..593
        /gene="A15"
        /number=1
        594..711
        /gene="A15"
        /number=1
        712..1013
        /gene="A15"
        /number=2
        1016
        /gene="A15"
        /note="start of hepta box"
        1025
        /gene="A15"
        /note="start of nona box"
        1466..1691
        /gene="A15"
        /note="alu repetitive sequence"
        1716..3062
        /gene="A16"
        1716
        /gene="A16"
        /note="start of dc box"
        1818..1866
        /gene="A16"
        /number=1
        1867..2061
        /gene="A16"
        /number=1
    exon
    intron

```

```

ACCESSION   AC003982
NID         g2769695
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS     Primates; Catarrhini; Homnidae; Homo.
TITLE       1 (bases 1 to 122302)
JOURNAL     Bradshaw, H, Wu, X and Ozersky, P.
REFERENCE   The sequence of H. sapiens PAC clone 166H1
AUTHORS     Unpublished (1998)
TITLE       2 (bases 1 to 122302)
JOURNAL     Waterston, R.
REFERENCE   Direct Submission
AUTHORS     Submitted (13-JAN-1998) Department of Genetics, Washington
TITLE       University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

MAPPING INFORMATION:
This clone was originally isolated in the laboratory of Professor
Graeme Bell, Howard Hughes Medical Institute and Departments of
Biochemistry and Molecular Biology, and Medicine, The University of
Chicago, Chicago, IL, USA. The clone was provided by the
laboratory of Dr. Roger Cox at The Wellcome Trust Centre For Human
Genetics, Oxford, UK. Some contig information was also obtained
from Yamagata et al., Nature 384:455-8 (1996).

SOURCE INFORMATION:
Clone 166H1 is from the first release of the human BAC library
CITB-978SK-B. The library contains cloned DNA from the male
fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl.
Acad. Sci. USA 89:8794-7 (1992); U-D. Kim et al., Genomics 34:213-8
(1996). This clone is available from Research Genetics, Inc.
(http://www.resgen.com).
VECTOR: pBeloBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is 278C19. The clone sequenced to
the right is 15F1. Actual start of this clone is at base position
1 of 166H1; actual end is at 122302 of 166H1.
Location/Qualifiers
1..122302
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="166H1"
/clone_id="CITB-978SK-B"
/map="12q"
1..238
/rpt_family="Alu"
239..265
/rpt_family="AT-rich"
268..565
/rpt_family="Alu"

```

```

repeat_region      10408. 10707
/rpt_family="Alu"
repeat_region      10713. 11014
/rpt_family="Alu"
repeat_region      11020. 11110
/rpt_family="Alu"
repeat_region      11146. 11386
/rpt_family="Alu"
repeat_region      11208. 11386
/rpt_family="Alu"
repeat_region      11391. 11515
/rpt_family="Alu"
repeat_region      11568. 11705
/rpt_family="Alu"
repeat_region      11737. 12032
/rpt_family="Alu"
repeat_region      12034. 12066
/rpt_family="7SLRNA"
repeat_region      12075. 12372
/rpt_family="Alu"
repeat_region      12448. 12753
/rpt_family="Alu"
repeat_region      12765. 12874
/rpt_family="t2"
repeat_region      12999. 13303
/rpt_family="Alu"
/rpt_family="Alu"
complement(13674. 13752)
/gene="WUSC:H.166H1.1"
/note="match to EST N80020 (NID:g1242721) za91a08.s1"
/note="match(13840. 14195)
/note="match to EST N80020 (NID:g1242721) za91a08.s1"
14196. 14229
/rpt_family="AT_rich"
repeat_region      14240. 14339
/rpt_family="t6"
repeat_region      14349. 14646
/rpt_family="Alu"
repeat_region      14652. 14940
/rpt_family="Alu"
repeat_region      14949. 15238
/rpt_family="Alu"
repeat_region      15807. 15937
/rpt_family="Alu"
repeat_region      15941. 16234
/rpt_family="Alu"
repeat_region      16236. 16408
/rpt_family="Alu"
repeat_region      16434. 16730
/rpt_family="Alu"
repeat_region      16745. 16791
/rpt_family="(TA)n"
repeat_region      16792. 17018

Note: remainder of annotations omitted.

Query Match      9.1%; Score 27; DB 26; Length 122302;
Best Local Similarity 79.5%; Pred. No. 3.77e-01;
Matches      35; Conservative      0; Mismatches      9; Indels      0; Gaps      0;

Db 106816 TATATAGTGTATATATAGTATGTGTATATAGTATAGTATATAA 106859
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 167 TACATATAGTGTGTATGTATGTATATATATATATGTAATGTGTGTTAA 210

RESULT 11
LOCUS      HSJ13817A 125787 bp DNA PRI 15-MAY-1996
DEFINITION Human DNA sequence from cosmid J138017, between markers DX86791 and
DX88038 on chromosome X contains Est CA repeat and an endogenous
retroviral like element.
ACCESSION      J772519
NID      g1301745
KEYWORDS      X.
SOURCE      human.

```

misc_feature	/note="MIT2A1 element fragment" complement(27005..32636)
repeat_region	/note="match: X89211 HBRV-L" 28155..29418 /note="L1 element fragment" complement(30111..30460) /note="match: 5' EST T74809 clone 85084" 33684..33749 /note="2 copies of 33 mer 96 % conserved" 34227..34430 /note="MIT2A1 element fragment" 34482..34615 /note="MIT2A1 element fragment" 35442..35585 /partial
repeat_region	/note="Alu repeat: matches 149..1 of consensus 35879..36028 /note="MIT1A element fragment" 36548..36697 /note="L1 element fragment" 36842..36881 /note="20 copies of 2 mer 100 % conserved" 36954..37079 /note="L1 element fragment" 37635..37757 /note="L1 element fragment" 38750..38851 /note="L1 element fragment" 40637..40807 /note="MER25 element fragment" 40700..41363 /note="L1 element fragment" 42523..42610 /note="L1 element fragment" 42692..43785 /note="L1 element fragment" 43013..43287 /note="MER25 element fragment" 44056..44255 /note="L1 element fragment" 44343..44652 /note="L1 element fragment" 44730..44918 /note="L1 element fragment" 44994..45742 /note="L1 element fragment" 46884..48969 /note="L1 element fragment" 50590..50847 /note="MER43 element fragment" 51224..51513 /note="MER25 element fragment" 51728..51760 /note="11 copies of 3 mer 100 % conserved" 51762..51917 /note="L1 element fragment" 51992..52203 /note="L1 element fragment" 52693..53288 /note="L1 element fragment" 53879..54079 /note="L1 element fragment" 54119..55495 /note="L1 element fragment" 55551..56370 /note="L1 element fragment" 56533..58176 /note="L1 element fragment" 58329..59175 /note="L1 element fragment" 59578..60572 /note="L1 element fragment" 60690..60988


```
repeat_region      /rpt_family="MER25"
                    65117. .69350
repeat_region      /rpt_family="L1"
                    69792. .72463
repeat_region      /rpt_family="L1"
                    complement(74228. .74550)
repeat_region      /rpt_family="THE1"
                    complement(74551. .76322)
repeat_region      /rpt_family="MSTAR"
                    74865. .75152
repeat_region      /rpt_family="ALU"
                    complement(76330. .76684)
repeat_region      /rpt_family="THE1"
                    77285. .77404
repeat_region      /rpt_family="MIR"
                    complement(81781. .81984)
repeat_region      /rpt_family="MER43"
                    83585. .87174
repeat_region      /rpt_family="L1"
                    84808. .85088
repeat_region      /rpt_family="ALU"
                    36211 c 35268 g 59355 t
BASE COUNT        66786 a 36211 c 35268 g 59355 t
ORIGIN
Query Match       9.1% Score 27; DB 26; Length 197620;
Best Local Similarity 72.9% Pred.No.3.77e-01;
Matches 43; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 22221 ATATATATATATATATATATATATATTTTAAATGCAACAGCTTCTACTTACA 22279
||| ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Cp 200 ATTACATATATACATACATACACACTATATGTATAGCAGCAATCTTGCCTACTACA 142

RESULT 13
LOCUS      AC005736 215441 bp DNA PRI 01-OCT-1998
DEFINITION Homo sapiens chromosome 16, BAC clone 462G18 (LANL), complete
sequence.
ACCESSION  AC005736
NID         93687215
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 215441)
            Rieke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
            Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
            Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
            White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.
            and Deaven,L.
            Sequencing of Human Chromosome 16p13.3
            Unpublished
            2 (bases 1 to 215441)
            Rieke,D.O.
            Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
            Unpublished
            3 (bases 1 to 215441)
            Rieke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
            Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
            Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
            White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.
            and Deaven,L.
            Direct Submission
            Submitted (01-OCT-1998) Center for Human Genome Studies, DOE Joint
            Genome Institute, Los Alamos National Laboratory, MS M888, Los
            Alamos, NM 87545, USA

FEATURES
            source
            Location/Qualifiers
                1..215441
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="462G18"
                /chromosome="16"

repeat_region      /map="16p13.3"
                    complement(417. .704)
repeat_region      /rpt_family="Alu"
                    complement(1146. .1543)
repeat_region      /rpt_family="THE1"
                    complement(2007. .2289)
repeat_region      /rpt_family="Alu"
                    2748. .2905
repeat_region      /rpt_family="Alu"
                    3386. .3646
repeat_region      /rpt_family="Alu"
                    complement(4300. .4382)
repeat_region      /rpt_family="Alu"
                    4676. .5059
repeat_region      /rpt_family="Alu"
                    6113. .6160
repeat_region      /note="(CCAT)12"
                    /rpt_type=tandem
                    /rpt_unit=CCAT
                    6236. .6478
repeat_region      /rpt_family="Alu"
                    6606. .6628
repeat_region      /note="(T)23"
                    /rpt_type=tandem
                    /rpt_unit=T
                    complement(6611. .6900)
repeat_region      /rpt_family="Alu"
                    complement(7270. .7413)
repeat_region      /rpt_family="MIR"
                    8232. .8353
repeat_region      /rpt_family="MLT1"
                    8384. .8384
repeat_region      /rpt_family="Alu"
                    complement(9006. .9599)
repeat_region      /rpt_family="Alu"
                    9305. .9329
repeat_region      /note="(T)25"
                    /rpt_type=tandem
                    /rpt_unit=T
                    complement(10517. .10798)
repeat_region      /rpt_family="Alu"
                    complement(11301. .11487)
repeat_region      /rpt_family="Alu"
                    11661. .11972
repeat_region      /rpt_family="Alu"
                    11974. .12005
repeat_region      /note="(TAA)8"
                    /rpt_type=tandem
                    /rpt_unit=TAA
                    complement(12492. .12584)
repeat_region      /rpt_family="Alu"
                    12682. .13107
repeat_region      /rpt_family="Alu"
                    complement(13106. .13222)
repeat_region      /rpt_family="Alu"
                    complement(13502. .13812)
repeat_region      /rpt_family="Alu"
                    13882. .14078
repeat_region      /rpt_family="Alu"
                    14061. .14086
repeat_region      /note="(A)26"
                    /rpt_type=tandem
                    /rpt_unit=A
                    complement(14288. .14424)
repeat_region      /rpt_family="Alu"
                    14450. .14692
repeat_region      /rpt_family="Alu"
                    complement(14718. .14812)
repeat_region      /rpt_family="Alu"
                    complement(14834. .15124)
repeat_region      /rpt_family="Alu"
                    15777. .16054
repeat_region      /rpt_family="Alu"
```

```

repeat_region      38846. .39409      /rpl_family="Alu"
repeat_region      40294. .40565      /rpl_family="Alu"
repeat_region      complement(40637) .40690      /rpl_family="MIR"
repeat_region      40752. .40826      /rpl_family="MIR"
repeat_region      complement(41089. .41164)      /rpl_family="Alu"
repeat_region      41229. .41478      /rpl_family="Alu"
repeat_region      complement(41530. .41853)      /rpl_family="THE1"
repeat_region      complement(41867. .42065)

...: remainder of annotations omitted.

Query Match      9.1%; Score 27; DB 27; Length 215441;
Best Local Similarity 75.0%; Pred. No. 3.77e-01;
Matches 48; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

Db 44309 ATATATATATGACTATGTTTAA-TACATATATATAGTGTGTGTATATATAT 44367
QY 137 ATATGTGAGTACGACAAAGTATTGCTGTACATATAGTGTATATATATATATGCT 196

Db 44368 ATAT 44371
QY 197 AAAT 200

RESULT 14
LOCUS      BTWCAM2      456 bp      mRNA      MAM      13-AUG-1997
DEFINITION      Bos taurus melanoma cell adhesion molecule (MCAM) mRNA, 3' UTR,
partial sequence.
ACCESSION      U089327
                g2323279
NID
KEYWORDS
SEGMENT
SOURCE
ORGANISM      Bos taurus
                Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
                Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae;
                Bovidae; Bovinae; Bos.
REFERENCE      1 (bases 1 to 456)
AUTHORS      Moore,S.S. and Byrne,K.
TITLE      An SSCP polymorphism at the bovine melanoma cell adhesion molecule
                (MCAM) locus
JOURNAL      Unpublished (1997)
REFERENCE      2 (bases 1 to 456)
AUTHORS      Moore,S.S. and Byrne,K.
TITLE      Direct Submission
JOURNAL      Submitted (11-FEB-1997) Tropical Agriculture, CSIRO, Carmody Road,
                St. Lucia, QLD 4072, Australia
FEATURES
source
1. .456
/organism="Bos taurus"
/note="located near NCAM locus"
/db_xref="taxon:9913"
/chromosome="15"
/tissue_type="skeletal muscle"
join(U089328:1..317,1..438)
/gene="MCAM"
<1..438
3' UTR
/gene="MCAM"
BASE COUNT      114 a      84 c      116 g      130 t      12 others
ORIGIN
Query Match      8.7%; Score 26; DB 19; Length 456;
Best Local Similarity 73.6%; Pred. No. 1.14e+00;
Matches 39; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```




Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mparch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 04:36:51 1998; Maspar time 305.11 Seconds

Tabular output not generated. 1256.814 Million cell updates/sec

Title: >US-08-530-112A-1219
Description: (1-174) from US08530112A.seq
Perfect Score: 174
N.A. Sequence: 1 GATCTAGCAGCGCAAGCTC.....AGCTGATTTGCAGTCAAA 174
Comp: CTRGACTCGTCGTTTCGAG.....TTTCACTTAACGTTCACTTT

Scoring table: TABLE default

Gap 6

Mismatch STD : Dbase 0; Query 0

Searched: 567134 segs, 1101898692 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

emb155
1:em_ba 2:em_fun 3:em_htg 4:em_hum1 5:em_hum2 6:em_in
7:em_com 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_vi
genbank107

Database:

15:gb_ba1 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_com 20:gb_ov
21:gb_pat 22:gb_ph 23:gb_pl 24:gb_pl2 25:gb_ptr1
26:gb_ptr2 27:gb_ptr3 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
32:gb_un 33:gb_vi

Statistics: Mean 9.569; Variance 6.424; scale 1.490

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB	ID	Description	Pred. No.
1	174	100.0	6513 27	AF016833	Homo sapiens maltase-g	2.65e+85
2	167	96.0	174 30	G25624	human STS EST51018.	5.21e+81
3	167	96.0	174 30	G27635	human STS SHGC-32830.	5.21e+81
4	36	20.7	10772 18	AF012089	Drosophila melanogaste	5.30e+05
5	34	19.5	10772 18	AF012089	Drosophila melanogaste	4.98e+04
6	28	16.1	7218 21	166494	Sequence 14 from paten	3.11e+01
7	28	16.1	38955 18	CEZC84	Caenorhabditis elegans	3.11e+01
8	28	16.1	42361 18	CEZC25D7	Caenorhabditis elegans	3.11e+01
9	26	14.9	215 21	128278	Sequence 5 from patent	2.37e+00
10	26	14.9	475 30	G27488	human STS SHGC-32310.	2.37e+00
11	26	14.9	10492 15	067572	Methanococcus jannasch	2.37e+00
12	26	14.9	103590 17	HS328E19	Human DNA sequence ***	2.37e+00
13	25	14.4	194 30	G38839	TA24 Plasmodium falci	6.36e+00

C	14	14.4	497 18	AF010535	Plasmodium falci	6.36e+00
C	15	14.4	732 23	MTBACOS2A	Banomulus mitochondri	6.36e+00
C	16	14.4	100754 17	ATC005489	*** SEQUENCING IN PROG	6.36e+00
C	17	14.4	105528 24	ATC00502392	Arabidopsis thaliana C	6.36e+00
C	18	14.4	139316 24	ATF7H19	Arabidopsis thaliana D	6.36e+00
C	19	14.4	164237 17	AC004838	*** SEQUENCING IN PROG	6.36e+00
C	20	14.4	216021 26	HUAC004787	Homo sapiens Chromosom	6.36e+00
C	21	13.8	356 18	DV95869	Dermacentor variabilis	1.68e+01
C	22	13.8	370 18	OF089262	Oxytricha fallax 57kD	1.68e+01
C	23	13.8	404 18	DAU39953	Disosmus areolator 16S	1.68e+01
C	24	13.8	12573 18	PFSC03006	Plasmodium falci	1.68e+01
C	25	13.8	34264 25	HSAC000060	Human BAC clone RG156N	1.68e+01
C	26	13.8	43310 26	AC004597	Homo sapiens chromosom	1.68e+01
C	27	13.8	74371 26	AC005369	Homo sapiens chromosom	1.68e+01
C	28	13.8	90271 17	AC004899	*** SEQUENCING IN PROG	1.68e+01
C	29	13.8	99083 18	CEY37A1B	Caenorhabditis elegans	1.68e+01
C	30	13.8	101715 23	ATF4D11	Arabidopsis thaliana D	1.68e+01
C	31	13.8	118357 27	HS1142C11	Human DNA sequence fro	1.68e+01
C	32	13.8	122940 17	AC004709	*** SEQUENCING IN PROG	1.68e+01
C	33	13.8	125686 27	HS503G16	Human DNA sequence fro	1.68e+01
C	34	13.8	209165 23	ATFCAS	Arabidopsis thaliana D	1.68e+01
C	35	13.8	236281 26	AC004673	Homo sapiens Xp22 BAC	1.68e+01
C	36	13.8	292720 17	CEY47H10	Caenorhabditis elegans	1.68e+01
C	37	13.2	316170 17	CEY37A1	Caenorhabditis elegans	1.68e+01
C	38	13.2	745 23	PINACT	P. contorta (lodgopole	4.33e+01
C	39	13.2	1740 15	MYC6KDSA	Mycoplasma hyopneumoni	4.33e+01
C	40	13.2	39056 25	HS096H1	Human DNA sequence fro	4.33e+01
C	41	13.2	92944 17	HS0242F8	Human DNA sequence ***	4.33e+01
C	42	13.2	128600 26	AC005192	Homo sapiens BAC clone	4.33e+01
C	43	13.2	197391 17	AC005140	*** SEQUENCING IN PROG	4.33e+01
C	44	13.2	206606 23	ATFCA1	Arabidopsis thaliana D	4.33e+01
C	45	13.2	217420 15	BSDB0017	Bacillus subtilis comp	4.33e+01

ALIGNMENTS

RESULT	LOCUS	1	AF016833	6513 bp	mrna	PRI	03-FEB-1998
DEFINITION	Homo sapiens maltase-glucoamylase mRNA, complete cds.						
ACCESSION	AF016833						
NID	92826520						
KEYWORDS	human.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;						
AUTHORS	Primates; Catarrhini; Hominiidae; Homo.						
TITLE	1 (bases 1 to 6513)						
JOURNAL	Nichols,B.L., Eldering,J., Avery,S., Hahn,D., Quaroni,A. and						
MEDLINE	Sterchl,E.						
REFERENCE	Human small intestinal maltase-glucoamylase cDNA cloning. Homology						
AUTHORS	J. Biol. Chem. 273 (5), 3076-3081 (1998)						
TITLE	2 (bases 1 to 6513)						
JOURNAL	Nichols,B.L., Eldering,J.A., Avery,S.E., Hahn,D., Quaroni,A. and						
FEATURES	Direct Submision						
source	Submitted (31-JUL-1997) Pediatrics, Baylor College of Medicine,						
	1100 Bates, CRC 10066, Houston, TX 77030-2600, USA						
	Location/Qualifiers						
	1..6513						
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
	/chromosome="7"						
	/tissue_type="small intestine"						
	/cell_type="enterocyte"						
	55..5628						
	/EC_number="3.2.1.20"						
	/EC_number="3.2.1.3"						
	/note="Brush border hydrolase; glucosyl hydrolase;						
	1,4-alpha-D-glucanohydrolase; has similarity to						
	sucrase-isomaltase"						
	/codon_start=1						

```
/product="malase-glucosylase"
/db_xref="PID:g2826521"
/translation="MARKKLKFTLLEIVSLVLLVLIISIVLILAKESIKSTAP
DDEGTGPDGTGTGPDGTGTGTHATTGTGPGGTGTGPGVSECVVLEIRINL
PDGPPTKATCDGRCNNPGAGNVSPCYSKNHSYVENLNTNATFARLKNLPS
SPVGSVNDVNLTAETQISNREHFKLTDNNFEVPHVSVFSNMAASLTYOYE
ISQOPSIKVTIRSNRNVLDSSIGPLLEFADQELQSTRLPSTNVIGLHVQOYRH
DMNKWPIFNDRITPNGNGLNLYGAOTFELCDEAGLSGVLKNSNMEVGLDPA
PALYRTIGLIDPEYVPLGNTPEQVOYELIGRPALEPSYMALGFEHLSREYGLDN
MRVVERNRAALPYDVHADIDMDRNPFTSDYFGEPEVEVNEHLNNGOKLVI
VDPALSNSSSSKPYGPDGSDPKIMVNSDQVTPGIGEMGOGTATPYNCAV
WMTRELFENOVFEDEIWDNMEVSNFVDSGSGSTNNLNPFPRIPLDGYLEK
TLCMDAVQHWKQYDHLNLYGSMAVATAAKTAEVFNKRSFILTRSTFGSGFEAH
WLGDNATWDDLKWSIPGVLEFNLFGIPMWGPDIQGLADTPELCKRMWOLGAYFP
SRHNGGYKQDOPASFGADSLINSRHLNRYTLPLYLTLFFRHRQGLVAFYF
LHFEYDNTGMDVHOOFIMWPGILLIPVLEDEGAERKAAVYDPAVWYFDEGSOVRR
KQVEMLEPDKDGLGHLRGVYIPPTQOQNTTLASRKNPILITILADENKAEGLFW
DDEKTDVANKVYLLEFSTYONRELVNISQTYKDPNNLAFNEIKILITEESANT
VKHNGVPSQTSPIVTDNSNLKVALITIDILLGEATVENSIRKIDREKIDCYPDENG
ASAENCTARCCIMEASNSGVPFCYFVNDLYSVDOYNHGAADISLSKSYANAF
PSTPVNLRDLDTYHKNEMPOFKIYDKNKREYVPLNIPSMPSSTPEQOLYVLK
KNPEGIEIRRSIGITILMSOLGFTSDMEIRSTRLPKXYIGFETEBHSYRDL
EMHTMGFSRDPGPKKNGYGVHPYMGLEOGSAHGVILLNSNAMDVTFQOPALY
YRTTGYLDYVYELGPTPELATOYTTELIGRPVAVPYWSIGFOLCRGYONDSIABL
YDENVAQIPIYVOYSDIDYERQDLFTLSKFAEPALINRMADMRVILILDPRI
SGNETQYPAFTRGVEDDVFIKYPNDQDIWGWKWPDPDVVNGSLDMSQVBLRYA
YVAFDPFFNSTAKMRKELEILNPNPERSLPGMWDNMEPSSFYNGAVSPGC
RDAALNPYPHLESRDRGLSKLQCSOOLPDGSLVOHVNHVNLWGSOTRPPY
EAVOEVGORGAVITRSTFPSSGWMAGHNDTAMDOLKSLIGMTESLFGISYT
GADICGFQDAEYEMCYRMWOLGAFYPSRHNNTIGTRODPSWMDVAFVMSITVO
TRECTLPLYLILHKAHTEGYTVVRPLHEFVSQVYTWIDISQLLGPAFLYSVLER
NARNVAVFPRARVYDYTGVDINAREGKTLPAFLDHLINLHVGGILPMQEPALNT
HLSROKPMGFRIALDEGTAGWGLFMDGOSIDYGLYLAFSQNTMOSHILF
NNYTGNNPLKGLYIEIMVGVSVPVTSASISVSGMVLTPSRNNDPTQVLSIDVTDRN
ISLHNTSLTWISTL"
103..165
misc_feature
/note="encodes membrane anchor region"
166..306
misc_feature
/note="encodes S-T rich stalk"
322..444
misc_feature
/note="encodes trefoil domain 1"
1633..1650
misc_feature
/note="encodes catalytic site 1"
2950..3042
misc_feature
/note="encodes trefoil domain 2"
4306..4323
misc_feature
/note="encodes catalytic site 2"
BASE COUNT 1747 a 1508 c 1562 g 1696 t
ORIGIN
Query Match 100.0%; Score 174; DB 27; Length 6513;
Best Local Similarity 100.0%; Pred. No. 2.65e-85;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
SOURCE human STS derived from sequences in dbEST and the Unigene
          collection.
ORGANISM Homo sapiens
           Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
           Vertebrata; Euteria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 174)
AUTHORS Hudson,T.
          Whitehead Institute/MIT Center for Genome Research; Physically
          Mapped STS
          Unpublished (1995)
JOURNAL
COMMENT
Contact: Thomas Hudson
          Whitehead Institute/MIT Center for Genome Research
          Whitehead Institute for Biomedical Research
          9 Cambridge Center, Cambridge MA 02142 USA
          Tel: 617 252 1900
          Fax: 617 252 1902
          Email: thudson@genome.wi.mit.edu
Primer A: GATCTGAGCAGCAAAAGCTC
Primer B: TTATTCTGGCATATAAATTGCAT
STS size: 147
PCR Profile:
  Presoak:
    Denaturation:
      Annealing: 36 degrees C
    Polymerization:
      PCR Cycles: 35
    Thermal Cycler:
      Template: 10 ng
      Primer: each 5 pM
      dNTPs: each 4 nM
      Tag Polymerase: 0.025 units/ul
      Total Vol: 20 ul
Buffer:
  MgCl2: 1.5 mM
  KCl: 50 mM
  Tris-HCl: 10 mM
  pH: 9.3
FEATURES
          source
          Location/Qualifiers
            1..174
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /map="643.0 CR from top of Chr7 linkage group"
            1..147
              STS
            1..20
              primer_bind
            68 a complement(124..147) 51 t 7 others
BASE COUNT 68 a 18 c 30 g 51 t
ORIGIN
Query Match 96.0%; Score 167; DB 30; Length 174;
Best Local Similarity 96.0%; Pred. No. 5.21e-81;
Matches 167; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```

```
Db 1 GATCTGAGCAGCAAAAGCTCAAAAGAGAGTTGGAGNTAAATAATTTATTTTTCGAG 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1 GATCTGAGCAGCAAAAGCTCAAAAGAGAGTTGGAGNTAAATAATTTATTTTTCGAG 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 TAGTGTGCTTTGAATGCTTAATCTTATTCTTAATGTAACACACATTTTCACATATA 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 61 TAGTGTGCTTTGAATGCTTAATCTTATTCTTAATGTAACACACATTTTCACATATA 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 AATATGCAATTATATGCGAGATAAATAAACAAGTGAATTTCGAAGNGAAA 174
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 121 AATATGCAATTATATGCGAGATAAATAAACAAGTGAATTTCGAAGNGAAA 174
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 3 G27635 174 bp DNA STS 28-JUN-1996
DEFINITION human STS EST51018.
ACCESSION G25624
NID 91347856
KEYWORDS STS sequence; primer; sequence tagged site.
```

DEFINITION human STS SHGC-32830.
 ACCESSION G27635
 NID 91396354
 KEYWORDS STS sequence: primer: sequence tagged site.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryotae: mitochondrial eukaryotes; Metazoa: Chordata: Vertebrata: Eutheria: Primates: Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 174)
 MYERS, R.M.
 REFERENCE 1 (bases 1 to 174)
 JOURNAL Unpublished (1996)
 COMMENT
 Contact: Richard M. Myers
 Stanford Human Genome Center (SHGC)
 Stanford University School of Medicine
 Department of Genetics, M-344, Stanford, CA 94305, USA
 Tel: 4157259687
 Fax: 4157259689
 Email: myerseshgc.stanford.edu
 Primer A: GATCTGACGAGCGAAGCTC
 Primer B: TTTTATCTGGCATATTAATTGCAT
 STS size: 147
 PCR Profile:
 Initial incubation: 94 degrees C for 90 seconds
 Denaturation: 94 degrees C for 15 seconds
 Annealing: 62 degrees C for 23 seconds
 Polymerization: 72 degrees C for 30 seconds
 PCR Cycles: 30
 Thermal Cycler: Perkin Elmer 9600
 Protocol:
 Template: 25 ng
 Primer: each 1 uM
 dNTPs: each 200 uM
 Taq Polymerase: 0.05 units/ul
 Total Vol: 10 ul
 Buffer:
 MgCl2: 2.5 mM
 KCl: 50 mM
 Tris-HCl: 20 mM
 pH: 8.3
 Prepared with primer pairs provided by Sandoz, derived from D20391
 -- Washington University/Merck EST sequence.
 FEATURES
 source
 1..174
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="7"
 STS
 primer_bind
 1..147
 complement(124..147)
 BASE COUNT 68 a 18 c 30 g 51 t 7 others
 ORIGIN
 Query Match 96.0%; Score 167; DB 30; Length 174;
 Best Local Similarity 96.0%; Pred. No. 5, 21e-81;
 Matches 167; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 Db 1 GATCGACGAGCGAAGCTCAAAAAGAGAGTTGGAGGNTAAATAATTTATTTTGCAG 60
 Oy 1 GATCGACGAGCGAAGCTCAAAAAGAGAGTTGGAGGNTAAATAATTTATTTTGCAG 60
 Db 61 TAGTGTCTTGAATGTGTAATNTTATNTTATGTATACACACATTCACATATAA 120
 Oy 61 TAGTGTCTTGAATGTGTAATNTTATNTTATGTATACACACATTCACATATAA 120
 Db 121 AATATGCAATTTATATGCCAGATAAANTTAAACAAGTGAATTTGCAAGNGAAA 174
 Oy 121 AATATGCAATTTATATGCCAGATAAANTTAAACAAGTGAATTTGCAAGNGAAA 174

RESULT 4
 LOCUS AF012089 10772 bp DNA INV 05-AUG-1997
 DEFINITION Drosophila melanogaster cysteine proteinase-1 (Cp1) gene, complete cds, and phenylalanyl tRNA synthetase gene, partial cds.
 ACCESSION AF012089
 NID 92305220
 KEYWORDS
 SOURCE
 ORGANISM
 Drosophila melanogaster
 Eukaryotae: mitochondrial eukaryotes; Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta: Pterygota: Diptera: Brachycera: Muscomorpha: Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 4546 to 4553)
 Gray, Y.H., Tanaka, M.M. and Sved, J.A.
 P-element-induced recombination in Drosophila melanogaster: hybrid element insertion
 Genetics 144 (4), 1601-1610 (1996)
 JOURNAL
 MEDLINE 97132596
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 3 (bases 1 to 10772)
 Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.
 Direct Submission
 Submitted (30-JUN-1997) School of Biological Sciences, University of Sydney, Biology A12, Sydney University, NSW 2006, Australia
 FEATURES
 source
 1..10772
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 join(872..1000,2310..2426,6476..6690,6751..7707)
 /gene="Cp1"
 /product="cysteine protease"
 872..7707
 /gene="Cp1"
 872..1000
 /gene="Cp1"
 /number=1
 1001..2309
 /gene="Cp1"
 2310..2426
 /gene="Cp1"
 /number=2
 join(2328..2426,6476..6690,6751..7462)
 /gene="Cp1"
 /codon_start=1
 /product="cysteine proteinase-1"
 /db_xref="PID:92305221"
 /translation="MRTAVILPLALLAVAAVSFAADVMEWHTEPKLEHRRKNYODET EERRKLTFFENKRIKAKHNRFAEGVSEFLAVKXADLHHEFRQLMGNFTTLK OLRAADSFKCVTFISPAHYTLPKSVDRKIGAVTAVKDSCHSCSWAFSTGLBEGO HFRKSGVLVSLSEQLVDCSTKYGNGCNGGLAMNARYIKDNGIDTEKSPYEALD DSCHEFNKGTGATRGFTDIPQGDCKMAEVAIVGPVSAIDASHSFQYSGVYN EPQCDQONLDHGVLVGFGTDESGEDYWLIVNSGTTWGDGFIKMLRNKNQCIAS ASSYPLV"
 2427..6475
 /gene="Cp1"
 4346..4353
 /gene="Cp1"
 /note="insertion site of p(Caspar)(50C)"
 /citation=[1]
 6476..6690
 /gene="Cp1"
 /number=3
 6691..6750
 /gene="Cp1"
 6751..7707
 /gene="Cp1"
 /number=4
 join(8110..9300,9370..>9532)
 intron
 misc-feature
 exon
 intron
 exon
 mRNA

[illegible][illegible]

NID	92724471
KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 7218)
TITLE	Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
JOURNAL	Recombinant fowlpox virus
FEATURES	Patent: US 5670367-A 14 23-SEP-1997;
source	Location/Qualifiers 1..7218 /organism="unknown"
BASE COUNT	1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN	
Query Match	16.1%; Score 28; DB 21; Length 7218;
Best Local Similarity	3.7%; Pred.No. 3,11e-01;
Matches	6; Conservative 90; Mismatches 68; Indels 0; Gaps 0;
Db	1059 CTGGCAATTTTTTTTTTTTTTTTTTTTTTTTTCGACATAATTATTTATGTGAAGA 110
Cp	CTTGCAAACTCAGTCTGTTATTTATTTATCGCATTAATAATGCATATTTATGTGAAGA 110
Db	1119 TTTTGGTTGTTATGATTTAACAATTAAGATTATTAACACTTCAAGCACACTACTGCCAAAAATA 50
Cp	TTGTGTTGTTATGATTTAACAATTAAGATTATTAACACTTCAAGCACACTACTGCCAAAAATA 50
Db	1179 TTTTGGTTGTTATGATTTAACAATTAAGATTATTAACACTTCAAGCACACTACTGCCAAAAATA 50
Cp	TTTGGTTGTTATGATTTAACAATTAAGATTATTAACACTTCAAGCACACTACTGCCAAAAATA 50
RESULT	7
LOCUS	CEZC84 38955 bp DNA INV 22-SEP-1998
DEFINITION	Caenorhabditis elegans cosmid ZC84, complete sequence.
ACCESSION	Z19157
NID	96927
KEYWORDS	Htg.
SOURCE	Caenorhabditis elegans.
ORGANISM	Caenorhabditis elegans
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae; Rhabditina; Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis 1 (bases 1 to 38955)
AUTHORS	Wilson,R.R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Dear,S., Du,Z., Durbin,R., Favello,A., Coulson,A., Craxton,M., Connel,M., Cosey,T., Cooper,J., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,W., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurtry,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Showkeen,R., Smaildon,N., Smith,A., Sonhammer,E., Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Sproat,J. and Wohlman,P.
TITLE	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL	Nature 368 (6466), 32-38 (1994)
MEDLINE	94150718
REFERENCE	2 (bases 1 to 38955)
AUTHORS	Thomas,K.
JOURNAL	Direct Submission
COMMENT	Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre Hinxton, Cambridge CB10 1RO, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: joesanger.ac.uk or rwenemate@wustl.edu Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information. For a graphical representation of this sequence and its analysis see:- http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence&object=ZC84 Current sequence finishing criteria for the C. elegans genome

sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. IMPORTANT: This sequence is not the entire insert of clone ZC84. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The start of this sequence (1..333) overlaps with the end of sequence Z19155. The end of this sequence (38670..38955) overlaps with the start of sequence Z19158.

Location/Qualifiers

1..38955

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="III"

/clone="ZC84"

1078..5093

/gene="ZC84.2"

join(1078..1399,1461..1652,1956..2632,2997..3173,3321..3314,33558..3713,3760..4226,4475..4558,4904..5093)

/gene="ZC84.2"

/note="similar to cyclic nucleotide gated olfactory channels"

/codon_start=1

/db_xref="PID:66928"

/db_xref="SWISS-PROT:O03611"

/translation="MSTAPADPPPTNSGSLAPTTNGISGPTTASAKFSLTFKVIILDHQCTHLPLHOLINFDNYEPFQFLRRKNQVHTTAQNEPFQKYNQNGSNVQQAATGQASDGSASIEVPKESYAVRIKRLIANTQDPDSDNYTCVTVAYIINLFLVIRQVFNLDIGSSQSLDFRTNGLSTNTOVECTYNNLTAKKEPTYSQYPLDGSKYMHFRLWLFDFLMDQVYLIDPLFNRMGVDOGLVAREAKYKAWQSKQVRIGISLIPLDYILGWPIPIYINRGRIPLRLRLRYKRRNCLERTETSSMPAFNRVYVYVYIVIIHMGACIYEWISGLGDVWYGHKLKQSLPDIDITLIRVYESYWSLTILTTIGTGEVSPVRNIEAFYTLDMCVLILFATINQSSMISMSAATTEQNKMDIKQIOMELRKYSKOLEPVIKWPDIYLTNKNOSLEDQYLVLPKIDIAETAMOVHEETLKRIVIFQDCEAGLAEVLVLAQOVSPDGFICKKDIGREMYIVRGLQVYDDGKKVFTVLDGSGVFELSLINAGSKGNRRITANRVSQYTDLPVLSITGLMNLAREYDPAKRLILAKGREILKKNLDENAPEQKTEETAEHLNNAVKLOTRMARLIVESHTEGKMKRIEMLEKLSCEPLFELFRKALAROKTWHGVSIDGGDS"TCGVDERRAPRLRQTKTIDLPTEGESLSLK"

complement(6212..10605)

/gene="ZC84.3"

complement(join(6212..6271,6321..6440,6492..7055,7510..7723,7811..8097,8175..8408,8462..8540,8594..8639,9158..9434,9483..9611,9660..10043,10519..10605))

/gene="ZC84.3"

/note="CDNA EST EMBL:C12747 comes from this gene; CDNA EST EMBL:C13723 comes from this gene; CDNA EST EMBL:C11504 comes from this gene; CDNA EST yk272g11.3 comes from this gene; CDNA EST yk272g11.5 comes from this gene"

/codon_start=1

/db_xref="PID:e300305"

/db_xref="PID:g1816483"

/db_xref="PID:g1816483"

/translation="MTSMOPKSGVYSKADTFKIFEDYKPVITSVADLRNKKFDVRIILSSDEMNKROTOLKTVRYSLVIEHKVEADVPMTAHVLQGLGFEIAYDLRQVIREALITCSFVYSKYGTETHSIGDILIVPMSOVNSTRIMASATLIEFVYQYTHQVTLISFSTSKDSQROLALVIVISKSDRIKQIMQICELIKSAINDDSKTSAGRAAFKAKIDEMSEEDALYIELDHSDKMRGGGASASMSVSEKSGIPIRSLSGSAKAMHINISAISEKKEGKILYPIADYSLNPIETKINGNLINLSASTMYLEIYSIFPHRSRLSEMLRLAKELFAKAAETLPNTKQIQGHLYLVLECFNAHOLVYCELMCDPIHLHARVAKVYLVLEIVTSLIDEYEPGASINAKELKTAIRKMLTADPRSLITLPHYKAKSCMFCVNVADFSALISDDSEQKNMHOVLQNGLENGISSNINATNSGATASRETNSVTFQKESTSGSEFPARRGGTGVNAGSINISNIALSLUEQSTSLRMEKVNINSVTVLPDPTLEKIQNVQDGLKRRSENNDEDESSISYVIMICGGGAGVWCCYAKLLNLRIEILSSKRSSEENKCLILGMCYCAQAKFDSIEMVACVDAVYNDAATALAVEDCRLTIAHPLPLSININIAKYILNOEDIDDERASIVYK

```

gene
/sequence="MTRLEFELPAEELNINIVDDITPTIIRAKYOSTSSIVKTVVYCLVAMVNRVEQRMPT
HFTLRKAMTNIIOYVNRATISTSLPRL"
complement(12747..18802)
/gene="ZC84.6"
complement(join(12747..12826,12887..13124,13172..13353,
13482..13976,14027..14212,14265..14561,14611..14807,
15064..15361,15411..15546,15592..15953,16318..16791,
16840..17013,17060..17182,17230..17598,17646..17973,
18041..18433,18484..18569,18796..18802))
/gene="ZC84.6"
/sequence="Similar to Serine protease inhibitor: cDNA EST
EMBL:D33965 comes from this gene; cDNA EST EMBL:D33199
comes from this gene"
/codon_start=1
/db_xref="PID:e349420"
/db_xref="PID:g2414433"
/db_xref="SPTREMBL:062504"
/translation="MFAQIDGLRDRKQRTGSGRVGACAFNIDCLAGMFCNGICGCLT
TYIAEGCYCKIDPGQCVYNEQCAVWPAFCIDTASGVTGCGENKVERATRDG
HACLDVLDANHTLITATCPLEPAGAGTALSIPKRPNDGDPVLCNTDSTINQD
EOAGDNGACLFPSNGQYIADQYDCEVSSIDLTSQSEKANGICCPNRAFTCVOP
TATGPNTEPRMWNYSITGMCQOFLMDSASGSGSHNNFTVEHSCFCIDTSRG
APYIHTSPLTEPPTGCSQTSACSHYECVSGTQMCPTVASGCGVGRPLD
SVNRGTYHSGVEKQGTATSTRFTYDQSGCSFTYLGAGNANFSLRDCGLYC
ARLQCDGNPLRIDVITSCANNDPSSHCKMDQAVCCPMMOTICTOPLRVGKDR
SVRRVYSAAATRECSFETGCGQNDNEETLVDCQTCRNAPAPPCQOQAYKQD
GKFTVCTNRQSSCCPANFECYFDGNMHCQCTKAFCTSLSPSKTGCGVSEFYHY
NPQDCESEFELGCDNSNTFASRAECENYCGVCGANGSPLEDSMGALQSCERD
GGCPSSHCYGLGPDMSYRCPTKTYICGLPQOGLSLGGILVTRRYFNITV
RKCSFVYNGCDGNPNFASINOCNFCMASACMDVYLLNPNTLPSCDDELONN
CPSFOCIYDLSLTDQVCCGATDMKVCYDNEKAYINSDGTYRELLINHNHNSCPDY
CFRNAIKRNFCCGSMKRNATCPTGAPRPDSTIGTPRTCTHAATSSCPDEFELSD
LKDAQYCCVSEICPHREXYFIDETSGMPSCTIGPHTVTPAGVQCAQIDGSMG
CCKGQPOLASDGGPPELIVMERNVATCDFNPNHNOCCPTFCQWISIRQRYQC
GSDPLTPMENDGCPTRQIAYTDPDKICTSALSIPADPAFCOFSONKQFOCC
MPSDQVMAFAITGTEAOACSMNGSACPGFSGVRKSGALCMADGIESEAOV
PYNVCILNRVLIGEOCEPTROCVGSHCIAMKSCPTGIEKOCOMAGSIREIKCT
DKOVAVDECFPLVTLGRSCVSAOQCGKCIDVDCDDESRYRGHCKDKOSRP
SLPMNVNFATGSPVHMKOIQPMSTSSAVITTTSSPOGHAELNNEBICMPK
OPLANGIAROCISGQPCVPGYSCWSPAKVYFCCANAKLWRYLYLVLDGEGAL
LEPATKAPVICTRYATCPGPGYCRSRPKTKRCKKROPLMMOSTIIOIKDITSRP
EPLMAHSAISAQVMAQSOYMHGRRRORPOPOKTKPKVKNKPTKTLPCPPNVYL
EMEVNRKTIKRCOATCPVMSBNGVCKIERRPVYTO"
complement(19472..25567)
/gene="ZC84.1"
complement(join(19472..19542,19619..19927,19975..20151,
20576..20692,20764..20961,21063..22028,22096..23043,
23091..23159,23211..23315,23362..23484,23559..23921,
23980..24232,24441..24847,25118..25190,25245..25300,
25552..25567))
/gene="ZC84.1"
/sequence="similar to Serine protease inhibitor, Kunitz type:
cDNA EST EMBL:D35965 comes from this gene; cDNA EST
EMBL:D33199 comes from this gene"
/codon_start=1
/db_xref="PID:e349419"
/db_xref="PID:g2414432"
/db_xref="SWISS-PROT:Q03610"
/translation="MEOSQFSGSGRGVDVQVNTDCLSGMYCATGLCGLSTYLRES
YCEKYNPNQPGCTVQVCEAVWPGSKRMDSSIGRCRCEEHVARETIDPWGICSL
RVSQGTAPIVYCALPBGAGFKISLNDSPMGAAPVCTGSSATIDPAGVAGG
GACMPSDSEYIDYDCTHTSPHILAEKFPQSNYAPNADYCCSRALAIQPOLPT
GMPSEPRMYNINVTGTCQDFMNDSCHTAKTISPNNKTIQGCESCRTQCRGSRQ
YTHESVHSERITSCAASVCAVSCAFQCTAIGSHLCCPFGSVCSKGGKRPDIQIR
TSIHGFGMYVYTKESIRFYDPTGRCQDFYVKGAGANGNNLSKHECMYARLQD
ERCSPLTIGEAORCONNAQCPSSHECKADQGVCPKQVTCAPLEIGICTNVKRY
WYAFTRCOMMEYTGCGQNDNPFIMDCNFCNAIPEPKCIQOGAYIYKQNTONEF
CSNGKCPNATVECTPFGSGWCCPTKATCSLNTDSGICQASTATKYXYPNPTONE
SFYNGCDGNSNPNFANDCESTCVGGCNGCTPLRDSGMYWCAQOOTSQSDSEH
CIPVAVGNSLINCRCPTRAYMGLPQOQTCANAYVQRYEINIVTSQCTPFGGDD
GNANFNLTIOQCNFCMSNACPAQNAVYDPNSQAMPINCEALANSQPSQACTQFNL
INNVCCGATSGVCEAEKAFIDITIDMSRECLINIEQSCPNYALCREFMOKNKKY
CSITKCTCQVGVKFLYKLRTPQIRCTIGRNDCCPDGFTCOSYLANAFQGFCTTANS

```

```

VCPDNALYLLDDSOQOPRACTOGTFVSCPNGYACRSILTSADGFCCKSGTGALPSPVT
DGGPNRYVTLVGEIATKQCDPRPNPNAPCSESEFTQMSLSNORQCCGSPAPRPOP
SDGCPNQLIARADAVRVCASAGANCPAGIFCOFSSNNOFCGCGSPRESVAF
VGMSEPEKCVAGQSCNCRGFAKORSLAGHVCTVKRVACESNEVLIIEGCHAVP
GSECLAVDOCCQVSCIEDACCECRPPLKAVAGFCOEDEIQQSSNOVNLCHNNAKLG
EACLTAVROCOENSCIGSCGCKRPELEKAGKCMTPVEKSAVSKSINNVCISKRG
MNPILIDEHEPRKHSICIRACPNAGYECFEPRTYLLNCPQKCTISRCPAGYECY
RNMYHCCSNANKRGLVAVCTGTSVSKSPSPDOECPRGNPLIYPTSSPYLCOG
KRQCPVGGCVQSSITSNNHICPSRHDPRQYQSESGLAGAYISRGPPAIPNPEGIL
VLTVRGVNRLEKRCERSCEPPQPIGQVCELNKAT"
26802..27381
/gene="ZC84.5"
join(26802..26998,27063..27149,27201..27381)
/gene="ZC84.5"

Note: remainder of annotations omitted.

Query Match      16.1%; Score 28; DB 18; Length 38955;
Best Local Similarity 65.9%; Pred. No. 3,11e-01;
Matches 58; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Db 38338 AGGTTTAAAGCAATCTTTTTCAGTAATGTTATTTCAAAATATAAAAAATTTCT 38397
||||| || ||||| ||||| ||||| || || || ||||| |||||
Qy 35 AGGTAAATAATATATATTTTTCAGTAGTGTGCTTGAAATGTAAATCTTATTCTA 94
||||| || || ||||| ||||| ||||| ||||| ||||| |||||
Qy 95 AGGTATACAAACACATTTTCAATATAAAA 122

RESULT 8
LOCUS CEC25D7 42361 bp DNA INV 21-SEP-1998
DEFINITION Caenorhabditis elegans cosmid C25D7, complete sequence.
ACCESSION Z81039
NID 91627604
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1
AUTHORS Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, J., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurtry, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Ritken, L., Roopra, A.,
Saunders, D., Showken, R., Smaildon, N., Smith, A., Sonnenhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Woldman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
Nature 368 (6466), 32-38 (1994)
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 42361)
AUTHORS Ainscough, R.
TITLE Direct Submission
COMMENT Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridgeshire CB10 1RO, England and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or tw@nematode.wustl.edu
coding sequences below are predicted from computer analysis, using
predictions from GeneFINDER (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:
http://web.sanger.ac.uk/cgi-
bin/tisplay/db-wormacc/class=Sequence&object=C25D7
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a

```


dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. This sequence is the entire insert of clone C25D7. The true left end of clone C01G10 is at 35450 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence 281061. The end of this sequence (35450..42361) overlaps with the start of sequence 281030.

FEATURES

source	location/Qualifiers
	1..42361
	/organism="Caenorhabditis elegans"
	/db_xref="taxon:6239"
	/chromosome="V"
gene	/clone="C25D7"
	/complement(3754..4270)
	/gene="C25D7.1"
CDS	/complement(join(3754..4107,4172..4270))
	/note="predicted using GeneFinder; cDNA EST EMBL:C13397 comes from this gene; cDNA EST EMBL:C11360 comes from this gene; cDNA EST yk279f9.3 comes from this gene; cDNA EST yk279f9.5 comes from this gene"
	/codon_start=1
	/db_xref="pid:e276084"
	/db_xref="pid:p1627608"
	/translation="MEKSKTSELSKKDEIDEANSESSKVPYLAIDPEEAKLPNAGGS EHVYNTSKRMALIKVRCGNALFRVEPTMHILEPNKCOLITIRMPGPIOKDAIYQ LQIENDVOPKTRAKADSAGCTPIPHLKIKIVAGASGROMSREYVDE"
gene	/complement(5248..6288)
	/gene="C25D7.2"
CDS	/complement(join(5248..5472,5534..5785,5835..6134,6205..6288))
	/note="cDNA EST yk364a11.3 comes from this gene; cDNA EST yk364a11.5 comes from this gene"
	/codon_start=1
	/db_xref="pid:e330171"
	/db_xref="pid:p16274870"
	/db_xref="sptrembl:017595"
	/translation="MISLVSLITASNTSVYLAAGSSKQDAASATKSSASQXSA KTRSGSKSGHLSGKPSKSKSKSVPLAVGARGAPKSESKLKRDKSKS SEKSRSSSKKCDKSAKCDLNKANKLCPVSVSDASIKNSGSEKSKALKLP NOSNTFPGOGTASVKTSLIRASPNKLTPATGVGVOTVIANNTSKRKAFKXTSON LYRVNPFVGFVEPDKLSIDVLRHNGVEXTDMIVLTLSNASEQNCARKVFEESDQPR ELIVPLVYN"
gene	/complement(6810..15244)
	/gene="sdc-3"
CDS	/complement(join(6810..7038,7124..7401,7602..7823,7867..7931,7981..8082,8135..8583,8909..9824,9870..10214,10655..13198,13521..13661,13839..14151,14206..14360,14408..14503,15037..15244))
	/note="sdc-3"
	/note="similar to zinc finger, C3H2 type; cDNA EST EMBL:D33689 comes from this gene; cDNA EST EMBL:D36575 comes from this gene; cDNA EST yk356g7.3 comes from this gene; cDNA EST yk214e7.3 comes from this gene; cDNA EST yk214e7.5 comes from this gene; cDNA EST yk297h11.5 comes from this gene; cDNA EST yk462a11.5 comes from this gene"
	/codon_start=1
	/db_xref="pid:e330172"
	/db_xref="pid:p16274871"
	/db_xref="sptrembl:017596"
	/translation="MSDEPIVLTREDEGPFNSCVSIAPVNDKSTIVELRIKAIAGK VFTQGLVVAARLYLDGELERPDLYTHVAADYDPFMPDPHRRMWEIPLPREP RLIDRLVOKIKMSNTYKQAKRRKRLRITHRKRGSEFLYLMHSEKNDVYKRY MAAYQPLPEKCIROVYREELNDEPSPSVVHHSKIRPATPDLLAKIAYLGKD FLVLTPEFRSOFRLRYKSKSPFRSNLEKWLKFKYKQDVGNIRDAAREVGL LORNVYHMTDPOCEODFMDVYRLQOMLENDIENITMLRENLSVTDDEENRSYV IMPENRASHIDVFTTDAEINLKPIVLELFNNMPLIEDIRVKECIDMIMLTVCQ
gene	
	/complement(join(24413..24531,24579..25231,25285..26649,26701..26799,26839..27047))
	/gene="C25D7.6"
CDS	/complement(join(24413..24531,24579..25231,25285..26649,26701..26799,26839..27047))
	/note="predicted using GeneFinder; Similarity to Xenopus DNA replication licensing factor MCM3 (SW:049739); cDNA EST EMBL:M75844 comes from this gene; cDNA EST EMBL:M89318 comes from this gene; cDNA EST EMBL:D27738 comes from this gene; cDNA EST EMBL:D27737 comes from this gene; cDNA EST EMBL:100522 comes from this gene; cDNA EST EMBL:101528
gene	
	/complement(18527..20451)
	/gene="C25D7.4"
CDS	/complement(join(18527..18736,18779..19421,19772..19855,20411..20451))
	/gene="C25D7.4"
	/codon_start=1
	/db_xref="pid:e330173"
	/db_xref="pid:p16274872"
	/db_xref="sptrembl:017597"
	/translation="MDYENGVDGFEEDMTASPSSELPITLVGEVINKLPIDRLI AKVSSKFOKILMYMPEKREIVNDIERVHILIDGNVLYQPTTGCYKRNHFS SLENDHYETAIIDLILIRKPKLELHLSVILIEILKKEERDLMTFLELKRREY INKTTISFTRMHCHEMKSLISLNNANNIKELINFHGLILLENYLAQLDEKNAKELC ASAVTFEFTYSLENLNFRKFEINLETFYADAVKKNILNLNNANNCYCEVSMQFV AHFEVAPRLFOPSPNTNGNDYFAFKCSGCVIIFLDYHLEIQRENDES"
gene	/complement(21896..24171)
	/gene="C25D7.5"
CDS	/join(21896..22064,22107..22210,22274..22436,22482..22648,23331..23540,23854..23919,23965..24171)
	/gene="C25D7.5"
	/codon_start=1
	/db_xref="pid:e330174"
	/db_xref="pid:p16274873"
	/db_xref="sptrembl:017598"
	/translation="MSFCOPDGSQDYVCCGTPELVLVYLNQKQSYENMNFYPROS HYDYSVIRSSKOLESDAFAKNOVILITPKTTTITVTRKQPVKSNVLYOMLTI PSLRSGNYSNLKILHTNSVLRDGAFCVFDVDTLPQKOKISLNLASYGPAKIRFV VYSSGPOPSGNLNLFPFOFIWADATMKDNVFKRREKHSITLSPGNSIIOSTP GPPNSITVYIRVDLMGTATAGALFPNGNDLVNDRSNYDIDILIERNALGKI KTRNIRETVNVLKKNDETHVRIELKPMPEPDTNRKAAVOANTPLVGDHRLCTGIQ YROEKAKRGKARMSKSTGVOTVSTKNQ"
gene	/complement(24413..27047)
	/gene="C25D7.6"
CDS	/complement(join(24413..24531,24579..25231,25285..26649,26701..26799,26839..27047))
	/note="predicted using GeneFinder; Similarity to Xenopus DNA replication licensing factor MCM3 (SW:049739); cDNA EST EMBL:M75844 comes from this gene; cDNA EST EMBL:M89318 comes from this gene; cDNA EST EMBL:D27738 comes from this gene; cDNA EST EMBL:D27737 comes from this gene; cDNA EST EMBL:100522 comes from this gene; cDNA EST EMBL:101528

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
2 (bases 1 to 10492)
Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D., Sutton,G.G., Blake,J.A., Fitzgerald,L.M., Clayton,R.A., Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J., Adams,M.D., Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G., Fuhmann,J.L., Nguyen,D.T., Utterback,T., Kelley,J.M., Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A., Roberts,K.M., Kaine,B.B., Borodovsky,M., Klenk,H.P., Fraser,C.M., Smith,H.O., Woese,C.R. and Venter,J.C.
Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*
Science 273 (5278), 1058-1073 (1996)
66337999

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
1.10492
/organism="Methanococcus jannaschii"
/db_xref="taxon:2150"
71.1549
/gene="MJ1319"
/note="similar to GB:142023 SP:P44849 PID:1005680
PID:1220824 PID:1204983 percent identity: 38.52;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="sodium-dependent transporter"
/db_xref="PID:g1591959"
/translation="MSYMERSESSNMGFLILASVSGAIGLGNIMRFQWYNTGGAGF
LIPYVALLCVGIPMLIEFAGHYTKRSAPLAEKLEKSGEWTFQAVISGFIISY
VYIIMCGLYLLIIVYIGYSPDNAPFFHNIISGVEDIGGVSGLVSLTAWG
YVALISAGIKNGLEKANKIMPFLLIILVNLNTLPGALTEGTEWLTLPDSKLF
NINWLSAFSQLFSLGFLIAVASYLPKKSDDLINAVVSLNCGSEPLAGAV
FGLUGMSYSTGPIPDKAVSEGILAFVFPKALISLPFASRFGFVFLVAFAGIS
SAVSIASVSALIDKESLSRKALVALALFIIISPIFTGGLYLDIIDHFAAGY
LPIAILEITAIIMFGDGKLEHVNKLSEIKLGYWVAGVSPILTLAVVELDA
SNVLTSGYGGYKTVYVPGALITPLAFVSVILQKKITKGW"
complement(2910.3305)
/gene="MJ1320"
complement(2910.3305)
/note="similar to GB:AE000782 percent identity: 35.04;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/db_xref="PID:g1591960"
/translation="MYKIYDPTNELLIVYVKKHINPYETERALNTKEETIILSPIKEE
LERLKLKRLKGRKELAVNLALAKIKYKLVDTYTADEAILNTAKENENYIVAIND
KELEKLENNIPVAVVROKKIFETFGW"
complement(3326.5467)
/gene="MJ1321"
complement(3326.5467)
/note="hypothetical protein; identified by Genemark;
putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJ1321"
/db_xref="PID:g1591961"
/translation="MYLNKVTYKINAKYKKEEIPKEVHFYRIKSVNENAFNRYRVN

gene
CDS
/gene="MJ1322"
complement(5486.8503)
/note="similar to PIR:A55094 percent identity: 27.01;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="putative NTPase"
/db_xref="PID:g1591962"
/translation="MSMLKEIRMANPKSHVNSRIKPEKGIYAIIGSGSKSIFEA
VFFALFGAGSNFYDTITKGSYVELDFEYVNNKRIIEYDGGAGAKLYNGK
PVATLISAVNKAVNELIGVDRMPLNSIYKOGELAKLSLSPSKLETVAKLLGIDE
FEKCYQKAGEIYKYEKRLERIEGLNKTENYERKELKNSOLEKNNKMETNDKN
KIKKEFEDIEKLEWENKTKLYEFINKLEERKALEKNOELILEYDLTVEAR
ETLNRHDEYERKSLVDEIRKIESRLKSHYDALIKRQLEIIGDLEKLEFI
NKSRYRDIIDNLTLNKLIDEIYERKIDLEELKLNNEIEKIEYKRCCEKE
YKEYLEDEERKAVENKLTLEYITLLOEKRSKKNINDETRINKLETRNIDESI
ENSLKEIEKKKLEINLOKRETELKKNKIGENSETIKRLKILDELAKEVGCPLKTP
IDENKRMELINOKTOLNKTLEYELEINKRIETIKRDLKLEKLEIDENKLTLEY
LEKQSOLEELKLNKYEQDEINKKISNVIINRQVDELLIEDIKSOLNKNRYNQ
YLSAVSYNSVDEGIRNRKIEINIVSGWNNKREKINRLEDEREINRKLKNE
LNKKEKLEIENRSLKFEKRYEGLTEKLEELKINDEGEIYNICNSKILADN
IKRKNKEDIEIYLNKILEYKLEINDEIRSYINOKLIDEINYEERKKIKEYEN
KROELDNVREKTEIETEGYELKKVDESKAKLEMSLEKREKLTAFVEYLDVVR
ITGRNGFOAYLEKTYVPLQKTLNFAESFDLPISFVELTKDFEVHAPNVLTIDN
LSGEOIAVALSLRIANALIGNVECIILDEPVTYLDENRRALAEIIFRVKSIPO
MIIITHRELEADVAVIINVKKDGNAVSKVING"
complement(8500.9600)
/gene="MJ1323"
complement(8500.9600)
/note="similar to SP:009683 percent identity: 28.42;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA repair protein RAD32 (rad32)"
/db_xref="PID:g1591963"
/translation="MMFVHIDNHLGYROYNLDRKDIYDSFKLCIKRILKIPDVV
LHSGDLFNDLRPRVPRKRAMQAFKRLKLENNIKVYIVAGNHEGPRRLSESPALLD
YKILIDGKVINVNGEIEIFIGTYHKKSKREMDLKNPSEKAKYKKKILMLHOG
INPYILDEYELHPDLPRKSYALHGIKRIETRNDGILASGSEITIIYREYDYG
KEGKGFYLVDESGNDLSDIKIDIECEPEYVNIKKKSGNEVNIERKKNPV
FGLIKRERKPFWDTLKDILINKALIVDEFDIMPDVNDIESLNTKELLVDYANOGI
DGDVLSTYKALNNENMKELDEYVNTKFRG"
9695.10399
/gene="MJ1324"
9695.10399
/note="similar to GB:U19201 SP:P32125 PID:304962 GB:U00096
PID:1790287 percent identity: 32.12; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="molybdopterin-guanine dinucleotide biosynthesis
protein B (mobB)"
/db_xref="PID:g1591964"
/translation="MRVIGVIGYKDSGKTTLLIEILKHSDDKIAVIKHTRKEDVNDK
GTDYRLSNAGAKITVLAIDSKVTFETRMDELNTLSVSDINIDFVIEEGKELKR
LNIPIKIVMLKDGSDILIDHRAMLIEPYNIIIDVIVYIKAVAPPMNINCGCY
NRTITVKAIVKGEAKRMDCVLAKGIITIVDGKTIIPAVFVSATVSTIKAMETLKG

Not I and Eco RI sites of the modified pT773 vector.
Library constructed by Bob Barstead."
/db_xref="taxon:9606"
/clone="IMAGE:1414501"
/clone_lib="Barstead spleen HPLR82"
/sex="male"
/dev_stage="adult, 17 years"
/lab_host="DH10B"
BASE COUNT 112 a 107 c 56 g 183 t
ORIGIN

Query Match 100.0%; Score 174; DB 13; Length 458;
Best Local Similarity 100.0%; Pred. No. 2,256-189;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 TTTCACATGCAATTCAGCTGTTTATTTATTTATTCGCAATTAATTCGATATTTATG 61
CP 174 TTTCACATGCAATTCAGCTGTTTATTTATTCGCAATTAATTCGATATTTATG 115
Db 62 TGAATGCGGTGTATACATAGAAATAGATTACATTTCAAGACACTACTGCA 121
CP 114 TGAATGCGGTGTATACATAGAAATAGATTACATTTCAAGACACTACTGCA 55
Db 122 AATATAATTTATTTACCTCCAACTCTCTTTGAGCTTTCCTGCTCAGATC 175
CP 54 AATATAATTTATTTACCTCCAACTCTCTTTGAGCTTTCCTGCTCAGATC 1

RESULT 2 AA894763 440 bp mRNA EST 09-JUN-1998
LOCUS OJ55a12.s1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1502206 3'
DEFINITION mRNA sequence.
ACCESSION AA894763
MID g3031164
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 440)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
www.Dio.liml.gov/db/rrp/image/image.html

Insert Length: 915 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amerisham
High quality sequence stop: 429.
Location/Qualifiers
1..440
/organism="Homo sapiens"
/note="Organ: kidney; Vector: pT773D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT773 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."
/db_xref="taxon:9606"

/clone="IMAGE:1502206"
/clone_lib="NCI_CGAP_Kid3"
/lab_host="DH10B"
BASE COUNT 109 a 105 c 53 g 173 t
ORIGIN

Query Match 98.3%; Score 171; DB 15; Length 440;
Best Local Similarity 100.0%; Pred. No. 2,356-185;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 CACTTGCAGAAATTCAGCTGTTTATTTATCTGGCATATAAATTCATATTTATGTCGA 60
CP 171 CACTTGCAGAAATTCAGCTGTTTATTTATCTGGCATATAAATTCATATTTATGTCGA 112
Db 61 AATGCGTTGTATACATTAGAAATAGATTACACATTTCAAGCAGCAGCTGCAAAA 120
CP 111 AATGCGTTGTATACATTAGAAATAGATTACACATTTCAAGCAGCAGCTGCAAAA 52
Db 121 TAAATTTATTTTAACTCCAACTCTCTTTGAGCTTTCCTGCTCAGATC 171
CP 51 TAAATTTATTTTAACTCCAACTCTCTTTGAGCTTTCCTGCTCAGATC 1

RESULT 3 HUMGS01365 174 bp mRNA EST 17-JUN-1996
LOCUS Human HL60 3'directed MboI cDNA, HUMGS01365, clone pml864, mRNA
DEFINITION sequence.
ACCESSION D20391
MID 9501487
KEYWORDS EST; EST(expressed sequence tag); HL60; gene signature(GS);
granulocyte; macrophage; monocyte; promyelocyte.
SOURCE Homo sapiens adult female promyelocyte cell_line:HL60 cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Eukaryota; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae;
Homo.
REFERENCE 1 (sites)
Muraoka, K., Matsubara, K., Fukushima, A., Yoshii, J. and Okubo, K.
Chromosomal assignments of 3'-directed partial cDNA sequences
representing novel genes expressed in granulocytoid cells
Genomics 23 (2), 379-389 (1994)
JOURNAL MEDLINE
REFERENCE 2 (sites)
95137584
Okubo, K., Itoh, K., Fukushima, A., Yoshii, J. and Matsubara, K.
Monitoring cell physiology by expression profiles and discovering
cell type-specific genes by compiled expression profiles
Genomics 30 (2), 178-186 (1995)
JOURNAL MEDLINE
REFERENCE 3 (bases 1 to 174)
96163870
Okubo, K., Fukushima, A., Yoshii, J., Niiyama, T., Kojima, Y.,
Yoshinari, H., Arimoto, J. and Matsubara, K.
Gene expression of human promyelocytic cell line HL60 before and
after induction of differentiation. A new application of
3'directed cDNA sequencing
Unpublished (1993)
JOURNAL MEDLINE
REFERENCE 4 (bases 1 to 174)
Okubo, K.
Direct Submission
Submitted (10-SEP-1993) to the DDBJ/EMBL/Genbank databases. Kousaku
Okubo, Osaka University, Institute for Molecular and Cellular Bio;
1-3, Yamada-oka, Suita, Osaka Pref. 565, Japan
(E-mail: kousaku@imcb.osaka-u.ac.jp, Tel: 06-877-5111(ex.3315),
Fax: 06-877-1922)
Submitted (10-Sep-1993) to DDBJ by:
Kousaku Okubo
Institute for Molecular and Cellular Biology
Osaka University
3-1, Yamadaoka
Suita, Osaka, 565
Japan
Phone: 06-877-5111
Fax : 06-875-1922.
Location/Qualifiers

Contact: Maria M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mousetest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:602436
vutative full length read
vector to vector length is 374
Seq primer: -28m3 rev2 ET from Amersham
High quality sequence stop: 21'.

```

/organism="Mus musculus"
/strain="FVB/N"
/Note-Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site.1: EcoRI. Site.2: NotI. 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTGCAACATCTGAAGGAGGAGGCGGCCCTTTTCTTTTCTTTTCTTTT
3']. double-stranded cDNA was ligated to Eco RI adaptors
(AATTCGATCCTTGG) digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Source irradiated bowel harvested 72 hours after
irradiation (1400 Gys). Library constructed by Bob
Harstead."

```

Query Match	16.1%	Score 28;	DB 13;	Length 302;
Best Local Similarity	68.1%;	Pred. No. 1.18e-06;		
Matches	62;	Conservative	0;	Mismatches 28;
			Indels 1;	Gaps 1.

D_b
 190 GATTTAAGTAGNAGSACTGATGTGCAACGAGCTGGATGTACAAATAAGTGCCTTA-AG 248
 ||| ||| ||| ||| ||| |||
 1 GATTCGACACCGCNAAGCTCAAAGAAGAGTTCGAGGTAAAATAATTATTATTTTSCAG 60

D_b
 249 AAGTATGCTTCAAGTGTATATAATCTTATT 279
 || | ||| ||| ||| ||| ||| |||
 61 TAGTGTCGTTGAANAATGTGTAAATCTATTT 91

Contact: Wilson RK
Washu-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAE Consortium (info@mae.llnl.gov) for further information
Insert length: 626 Std Error: 0.00
Seq primer: -40m3 fwd. from Amersham
High quality sequence stop: 269.

/organism="Homo sapiens"
 /note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I;
 Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
 oligo(dT) primer [5',
 AACGCGAAGATTCGGCGCCGCTTTT TTTT TTTT TTTT TTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by M. Patricia Bonaldo."

every Match	15.58;	Score 27;	DB 6;	Length 385;
1st Local Similarity	82.98;	Pred. No. 1.01e-05;		
Matches	34;	Mismatches	7;	Indels 0;
				Gaps 0;

DB	Qy	RESULT	LOCUS	DEFINITION	ACCESSION	NTD
61	112	8	R45560	422 bp	EST	22-MAY-1995
101	152		y444bD8.s1	Homo sapiens CDNA clone 35260 3'	similar to contains	
			MER32	repetitive element ;		
			R45560			
			9823774			

LOCUS AA116050 420 bp mRNA EST 23-DEC-1997
 DEFINITION clone 531863 5', mRNA sequence.
 ACCESSION AA116050
 NID 91671099
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 420)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisan, E.,
 Watson, R., Williamson, A., Woldmann, P., and Wilson, R.
 WASHU-Merck EST Project
 TITLE Unpublished (1995)
 JOURNAL COMMENT
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 WARNING: There is evidence that suggests that the 384-well parent
 plate of this clone contains both human and mouse derived clones.
 Thus, the origin of this clone is uncertain. This caution should be
 kept in mind should you use this clone.
 This clone is available royalty-free through LNL: contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1484 Std Error: 0.00
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 390.
 Location/Qualifiers
 1..420
 /organism="Homo sapiens"
 /note="Vector: pBluescript SK-. Site_1: EcoRI; Site_2:
 XhoI; Cloned unidirectionally. Primer: Oligo dt. NT2
 cells (Ntera-2/cl.D1) induced with Retinoic Acid for 24
 hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5'
 adaptor sequence: 5' CTCGACGTTTCTTTTCTTTTCTTTT 3' adaptor
 sequence: 5' CTCGACGTTTCTTTTCTTTTCTTTT 3'."
 /db_xref="GDB:3921375"
 /db_xref="taxon:9606"
 /clone_id="531863"
 /clone_lib="Stratagene neuroepithelium (#937231)"
 /dev_stage="Ntera-2/RA neuroepithelial cells"
 /lab_host="SOLR (kanamycin resistant)"
 BASE COUNT 109 a 94 c 84 g 131 t 2 others
 ORIGIN
 Query Match 14.4%; Score 25; DB 11; Length 420;
 Best Local Similarity 68.2%; Pred. No. 6.63e-04;
 Matches 45; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
 Db 247 GTTTATGTTATGCGGTTCTTTGTTATTTTGTGTTGTTGTTTAAACAT 306
 ||||||| ||||| ||| ||| ||||| ||||| ||| ||| |||||
 Cp 154 GTTTATTTTATCTGCATTAATGCAATATTTTATGCAATGCGTTTACAT 95
 Db 307 TCATAA 312
 | |||
 Cp 94 TAGAAA 89
 RESULT 14
 LOCUS AA0201229 427 bp DNA GSS 16-SEP-1998
 DEFINITION RBC111-45118.TJ RBC111 Homo sapiens genomic clone R-45118, genomic
 survey sequence.
 ACCESSION AA0201229
 NID 93613428
 KEYWORDS GSS.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 427)
 AUTHORS Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
 Berry, K., Granger, D., Suh, E., White, C., de Jong, P., and Venter, J.C.
 TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
 JOURNAL Unpublished (1998)
 COMMENT Other_GSSs: RBC111-45118.TK
 CONTACT: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 For clone availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 Research Genetics (info@resgen.com). BAC end search page:
 http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
 Class: BAC ends.
 Location/Qualifiers
 1..427
 /organism="Homo sapiens"
 /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
 RBC111 Human Male BAC Library"
 /db_xref="taxon:9606"
 /clone="R-45118"
 /clone_lib="RBC111"
 /sex="Male"
 /cell_type="lymphocytes"
 BASE COUNT 106 a 92 c 68 g 161 t
 ORIGIN
 Query Match 14.4%; Score 25; DB 28; Length 427;
 Best Local Similarity 72.7%; Pred. No. 6.63e-04;
 Matches 40; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
 Db 9 AATTCACCTTCTTTAAATAATGATTTATTAATGCAATCTTATCTTAAT 63
 ||||||| ||||| ||| ||| ||||| ||||| ||| ||| |||||
 Cp 163 AATTCACCTTCTTTATTTATTCGCAATTAATGCAATTTTATATGTAAT 109
 RESULT 15
 LOCUS AA701787 552 bp mRNA EST 25-JUN-1998
 DEFINITION PmIFG_1413 Douglas-fir (Pm.1000) cDNA library Pseudotsuga menziesii
 (Y12285), mRNA sequence.
 ACCESSION AA701787
 NID 93253367
 KEYWORDS EST.
 ORGANISM Douglas fir.
 Pseudotsuga menziesii
 Eukaryote; Viridiplantae; Charophyta/Embryophyta group;
 Embryophyta; Tracheophyta; seed plants; Coniferopsida; Coniferales;
 Pinaceae; Pseudotsuga.
 REFERENCE 1 (bases 1 to 552)
 AUTHORS Jermstad, K.D., Bassoni, D.L., Kinlaw, C.S., and Neale, D.B.
 TITLE Partial DNA sequencing of Douglas-fir cDNAs used for RFLP mapping
 JOURNAL Unpublished (1997)
 COMMENT
 CONTACT: Jermstad, KD
 Miroy Molecular Genetics Laboratory
 Institute of Forest Genetics, USDA Forest Service
 2480 Carson Road, Placerville, CA, 95667, USA
 Tel: 530-622-1225
 Fax: 530-622-2633
 Email: kdjes27w07.pswfs.gov
 sequence length = 552 bases.
 Insert length: 700 Std Error: 0.00
 Seq primer: T7.
 Location/Qualifiers
 1..552
 FEATURES
 SOURCE

```

/organism="Pseudotsuga menziesii"
/note="Vector: Uni-Zap XR lambda (Stratagene); Site_1:
BamHI; Site_2: EcoRI; variety = menziesii; mRNA was
isolated from new growth needle tissue and cloned
uni-directionally into Uni-Zap XR with the ZAP-cDNA
Synthesis Kit (Stratagene). The phagemid Bluescript SK
(-) was excised from the phage vector using the in vivo
excision protocol provided by Stratagene. The nucleotide
sequence of the 3' end was obtained by automated
sequencing at the Recombinant DNA/Protein Resource
Facility, Oklahoma State University, Oklahoma, USA."
/db_xref="taxon:3357"
/clone_id="Douglas-fir (Pm_1000) cDNA library"
/sex="hermaphrodite"
/tissue_type="needle tissue"
/dev_stage="spring growth"

```

```

BASE COUNT      182 a      79 c      120 g      127 t      44 others
ORIGIN

```

```

Query Match      14.4%; Score 25; DB 15; Length 552;
Best Local Similarity 78.6%; Pred. No. 6.63e-04;
Matches 33; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

```

Db      490 CATAAATATAGCAATTATATTCNNAAAAAAAAAAAAA 531
      ||||| || ||||| ||||| | ||||| ||||| ||
QY      115 CATAAATATGCAATTATATGCCAGATAAAATAAACAA 156

```

```

Search completed: Tue Dec 22 04:47:19 1998
Job time : 294 secs.

```


SQ	Sequence	174 BP;	70 A;	20 C;	30 G;	54 T;
	Query Match	100.0%;	Score 174;	DB 19;	Length 174;	
	Best Local Similarity	100.0%;	Pred. No. 2,99e-78;			
	Matches	174; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
Db	1 gatctggcgcggaagaagccataaaaggacttggaggttaaaaaataattatlttgcaq	60				
OY	1 GATTGAGCAGGCACCAACTCAAAGAAGCTTTGGAGGTAAAAATATATTATTTTTTGGCG	60				
Db	61 taagtgttcctttgaatatlytgtlaaatcttatlttcaatgylatacacaccattccaataa	120				
OY	61 TAGTGTCCTTTGAATATGTATAATCTTAATTTCAAAAGTATACACCACCATTTCCACATAA	120				
Db	121 aaatgccattatatacccagataaaaataaacaaagtagattgcaaagttaa	174				
OY	121 AAATATGCAATTTATATATGCCAGATAAAATAAACAAAGTAGATTGCCAAGTGAA	174				
RESULT	2					
ID	OS1746 standard; cDNA:	91 BP.				
AC	OS1746: (first entry)					
DE	31-MAY-1994 (first entry)					
Dn	Oligonucleotide probe MK14-A					
KW	Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;					
ss.						
OS	Synthetic.					
PN	EP-571911-A.					
PD	01-DEC-1993.					
PR	24-MAY-1993; 108325.					
PA	(BECT) BECTON DICKINSON CO.					
PI	Shank DD, Spears PA;					
DR	WPI: 93-378844/8.					
PT	New oligo:nucleotide probes specific for Mycobacteria - used for detection and amplification of Mycobacteriea nucleic acid in samples					
PT	Claim 3, Page 14, 23pp: English.					
CC	Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14 CC (O51735). It hybridized to all spp. of mycobacteria tested, but cross reacted to a few non-mycobacterial spp. The probe may be useful as an initial screen for mycobacterial infection.					
CC	See also O51735-45 and O51747-59.					
CQ	Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;					
Query Match	19.0%; Score 33; DB 9; Length 91;					
Best Local Similarity	5.9%; Pred. NO. 8.0le-05;					
Matches	3; Conservative 39; Mismatches 9; Indels 0; Gaps 0;					
Db	13 vhsyvvhvshvshbvhhvvbhvsvvvhvvhvvbhvyhsvctc 63					
Cp	77 CATTTCAAGCACACTACTGCAAAAATAATATTATTTAACCTCCAACATC 27					
RESULT	3					
ID	N81164 standard; DNA:	204 BP.				
AC	N81164;					
DT	08-NOV-1990 (first entry)					
DM	Base substituted E.coli beta-galactosidase alpha-fragment.					
KM	E.coli beta galactosidase alpha-fragment; base substitutions: ss.					
OS	Escherichia coli.					
FH	key Location/Qualifiers					
FT	misc_feature 19..69					
FT	/tag= a					
FT	/function=multiple cloning site					
FT	primer_bind 187..204					
FN	EP-285123-A.					
PD	05-MAY-1988.					
PF	30-MAR-1988; 105163.					
PR	03-APR-1987; US-034819.					
PA	(SUSO) SUDOMEN SOKERI OY,					
PI	Leltovaara P., Knowles J., KoivulaA., Bamford J., Reinkenainen T;					

```

DR      WPI, 88.279927/40.
PT      Introducing random point mutations into nucleic acids -
PT      by prep'n of single stranded template, annealing a primer, elongation,
PT      misincorporation, completion of molecules and screening.
PS      Disclosure: P; English.
CC      Random point mutations were introduced into the alpha fragment of
CC      E.coli beta galactosidase. The wild type sequence was obtained as a
CC      single stranded template and an oligonucleotide was hybridised to
CC      it to generate a pop'n of DNA molecules which terminate at all
CC      possible nucleotide positions within a specified region. The
CC      variable 3' ends generated in this way are used as primers for
CC      reverse transcriptase. Nucleotides are misincorporated by the
CC      transcriptase and the molecules are completed to forms that can be
CC      amplified and then expressed in a suitable host-vector system.
CC      The sequence covers then all 176 diff base substitutions, most of which
CC      occurred singularly in any given mutant.
CC      See also P80575.
SO      Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match      19.0%; Score 33; DB 1; Length 204;
Best Local Similarity 9.0%; Pred. No. 8.01e-05;
Matches 9; Conservative 53; Mismatches 37; Indels 1; Gaps 1.

Db      92 hhyrmrmdhnyrdynirsdaaayccyrtsyvdccycaachbhdhvyvbbbyvnhnmc 151
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Oy      68 CTTTGAAATGCTGAATCTTATTTCCTATGATGATCAACACACATTTCACATATAAAATATGC 127

Db      152 nccbnhb-vehvnbhnmhwayrbharrrdvncvch 190
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Oy      128 AATTATATGCGCAGATATAAATAAACAAGCAATTTGCA 167

```

ID	RESULT	4
AC	Q51746 standard; cDNA; 91 BP.	
AD	Q51746;	
DE	31-MAY-1994 (first entry)	
KM	Oligonucleotide probe MK14-A	
KW	Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;	
OS	Synthetic.	
PN	EP-571911-A.	
PD	01-DEC-1993.	
PF	24-MAY-1993; 108325.	
PR	26-MAY-1992; US-889651.	
PI	(BECT) BECTON DICKINSON CO.	
PA	Shank DD; Spears PA;	
DR	WPI; 93-378844/48;	
PT	New oligo:nucleotide probes specific for Mycobacteria - used for	
PT	detection and amplification of Mycobacteria nucleic acid in	
PT	samples	
PS	Claim 3; Page 14; 23pp: English.	
CC	Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14	
CC	(Q51735). It hybridized to all spp. of mycobacteria tested, but	
CC	cross reacted to a few non-mycobacterial spp. The probe may	
CC	be useful as an initial screen for mycobacterial infection.	
CC	See also Q51735-45 and Q51747-59.	
SO	Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;	
Query Match	18.4%; Score 32; DB 9; Length 91;	
Best Local Similarity	0.0%; Pred. No. 2.27e-04;	
Matches	0; Conservative 38; Mismatches 6; Indels 0; Gaps 0;	
DB	13 vhsyvvvvhvshhsbvhhvhhvsvvvvhhvhhvhhvhyhv 56	
QY	113 CACATTAATAATGCAATTATATGCCAGATAAATAAACAA 156	
RESULT	5	
ID	N81164 standard; DNA; 204 BP.	
AC	N81164;	
DT	08-NOV-1990 (first entry)	
DE	Base substituted E.coli beta-galactosidase alpha-fragment.	
DE	E.coli beta galactosidase alpha-fragment; base substitutions; ss.	


```

OS Escherichia coli. Location/Qualifiers
FH Key 19..69
FT misc-feature /tag= a
FT primer_bind /function= multiple cloning site
FT 187..204
FT /*tag= b
PN EP-285123-A.
PD 05-MAY-1988.
PR 03-MAR-1988: 105163.
PR (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
PI WPI: 88-27992/40.
PT Introducing random point mutations into nucleic acids -
PT by prep of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure: P: English.
CC Random point mutations were introduced into the alpha fragment of
CC E. coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a pop of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP: 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 16.7%; Score 29; DB 1; Length 204;
Best Local Similarity 10.5%; Pred. No. 4,93e-03;
Matches 11; Conservative 49; Mismatches 45; Indels 0; Gaps 0;

DB 87 mtctthyrmbvnyrdynrdaaawccyrrsvkydcacnchddhyvbbvnyh 146
CP 131 AATTGCATTTTATGTAATAATGTGTGATACATTTAGAGATTATACCATTTTC 72
DB 147 nhnnccebnhcnvbnhnrwayrvhrdarddhvccvchc 191
CP 71 AAGCACACTACTGCAAAATAATATTTTAACTCCAAACTC 27

RESULT 6
ID N50023 standard; DNA: 501 BP.
AC N50023:
CP 04-SEP-1991 (first entry)
DE Sequence encoding new modified human beta interferon polypeptides
DE IFN $\gamma$  415.
DE Antiviral; cell growth regulator; immune system regulator;
DE antiproliferative; ss.
DE Homo sapiens.
FH Key Location/Qualifiers
FT cds 1..501
FT /*tag= a
FT EP-163993-A.
FT 11-DEC-1985.
FT 17-MAY-1985: 105750.
FT 17-MAY-1984: GB-012564.
FT (SEAR) SEARLE G D & CO.
FT Bell LD, Boseley PG, Porter AG;
FT WPI: 85-311944/50.
FT P-PSDB: P50022.
FT New modified human beta interferon polypeptide(s) - prepd. by
FT plasmid transformed bacteria, with improved antiviral,
FT anti-proliferative and immune regulating actions
FT Claim 28: Chart 2a, page 32; 71pp; English.
CC Compared with interferon beta prepd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index; improved stability against microbial

```

```

CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixts.
CC Sequence 501 BP: 107 A; 31 C; 69 G; 80 T;

Query Match 14.4%; Score 25; DB 3; Length 501;
Best Local Similarity 28.3%; Pred. No. 2,60e-01;
Matches 17; Conservative 22; Mismatches 21; Indels 0; Gaps 0;

DB 25 ytbcarngdmwnnaayltycarwnccaraarylbtgtgcarytbnmygmgdytb 84
CP 60 CTCGCAAAATAATTTATTTTAACTCCAAACTCTCTTTGACCTTGCTGCTGAGATC 1

RESULT 7
ID T14325 standard; DNA: 70 BP.
AC T14325:
DE 16-JUN-1997 (first entry)
DE Conjugate formed by labelling and sorting of nucleotide molecules.
DE Labelling; sorting; sequencing; tag; tagging; ss.
DE Synthetic.
FH Key Location/Qualifiers
FT misc-feature 65
FT /*tag= a
FT /note= "This N represents the amplified cDNA
FT sequence"
FT PN W09612039-A1.
FT PD 25-APR-1996.
FT PF 12-OCT-1995: U12678.
FT PR 13-OCT-1994: US-322348.
FT PR 19-DEC-1994: US-359295.
FT PA (LYNX-) LYNX THERAPEUTICS INC.
FT PI Brenner S;
FT WPI: 96-222023/22.
PT Labelling and sorting mols. using oligo:nucleotide tags - useful in
PT large-scale parallel operations, e.g. DNA sequencing and mRNA
PT fingerprinting
PS Disclosure: Page 20; 71pp; English.
CC Determining the nucleotide sequence (I) of a target polynucleotide
CC (T) comprises: (a) generating from T a plurality of fragments that
CC cover T; (b) attaching an oligonucleotide tag from a repertoire of
CC tags, to each fragment such that all the same fragments have the
CC same tag, and all different fragments have different tags; (c)
CC sorting the fragments by specifically hybridising the tags with
CC their respective tag complements; (d) determining (I) of a portion
CC of each of the fragments; and (e) determining (I) of T by collating
CC the sequences of the fragments. The tagging system can be used with
CC single base sequencing methods to sequence polynucleotides up to
CC several kilobases in length. The tagging system permits many
CC thousands of fragments of a target polynucleotide to be sorted onto
CC one or more solid phase supports and sequenced simultaneously.
CC A primer which is initially used to reverse transcribe mRNA is
CC described in T14322. If the amplified product is then needed to be
CC attached to a solid phase support a sequence like the one given in
CC T14323 may be used. The mRNA would then be removed and the second
CC strand of cDNA produced using a primer with a similar form to that
CC described in T14324. After restriction enzyme digestion, the
CC conjugate would have a formula similar to this.
SQ Sequence 70 BP: 2 A; 12 C; 3 G; 19 T;

Query Match 13.2%; Score 23; DB 24; Length 70;
Best Local Similarity 18.6%; Pred. No. 1,76e+00;
Matches 8; Conservative 25; Mismatches 10; Indels 0; Gaps 0;

DB 1 rggaccacwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 43
CP 64 ACTACTGCATAAATAATATTTTAACTCCAAACTCTCTT 22

RESULT 8
ID N50033 standard; DNA: 501 BP.
AC N50033:
DE 04-SEP-1991 (first entry)
DE Sequence encoding new modified human beta interferon polypeptides

```

```

DE IFNX 456.
KW Antiviral; cell growth regulator; immune system regulator;
OS Homo sapiens.
FH Key 1.501
FT cds /tag= a
PN EP-163993-A.
PD 11-DEC-1985.
PE 17-MAY-1985; 105750.
PA (SEAR ) SEARLE G D & CO.
PI Bell LD, Boseley PG, Porter AG;
P-PSDB; P50032.
DR New modified human beta interferon polypeptide(s) - prepd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PS Claim 28; Chart 2k, page 42; 71pp; English.
CC Compared with interferon beta prepd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index; improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixts.
SQ Sequence 501 BP; 111 A; 31 C; 68 G; 80 T;

Query Match 13.2% Score 23; DB 3; Length 501;
Best Local Similarity 34.1% Pred. No. 1.76e+00;
Matches 14; Conservative 16; Mismatches 11; Indels 0; Gaps 0;

Db 25 ytbcaragdmwnmaaytlycartgycaraatybybtg 65
CP 60 CTGCAAAATAATATTATTACCTCCAACTCTCTTTG 20

RESULT 9
ID N50028 standard; DNA; 501 BP.
AC N50028;
DE 04-SEP-1991 (first entry)
DE Sequence encoding new modified human beta interferon polypeptides
DE IFNX 445.
KW Antiviral; cell growth regulator; immune system regulator;
OS Homo sapiens.
FH Key 1.501
FT cds Location/Qualifiers
FT EP-163993-A. /tag= a
PN 11-DEC-1985.
PD 17-MAY-1985; 105750.
PE 17-MAY-1984; GB-012564.
PA (SEAR ) SEARLE G D & CO.
PI Bell LD, Boseley PG, Porter AG;
P-PSDB; P50027.
DR New modified human beta interferon polypeptide(s) - prepd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PS Claim 28; Chart 2f, page 37; 71pp; English.
CC Compared with interferon beta prepd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index; improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixts.
SQ Sequence 501 BP; 114 A; 31 C; 68 G; 79 T;

Query Match 13.2% Score 23; DB 3; Length 501;
Best Local Similarity 39.0% Pred. No. 1.76e+00;
Matches 16; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

Db 25 ytbcaragdmwnmaaytlycartgycaraatybybtg 65

```

```

CP 60 CTGCAAAATAATATTATTACCTCCAACTCTCTTTG 20

:|:|:| : : | :|:| :|:|:|:|:|

RESULT 10
ID 076371 standard; DNA; 2676 BP.
AC 076371.
DE 23-JUN-1995 (first entry)
DE Mouse growth differentiation factor-8 DNA.
KW Growth differentiation factor-8; GDF-8; cell proliferation;
KW adipocyte; obesity; transforming growth factor-beta; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT cds 104..1231
FT /tag= a
PN WO9421681-A.
PD 29-SEP-1994.
PE 18-MAR-1994; U03019.
PR 19-MAR-1993; US-033923.
PA (UYO ) UNIV JOHNS HOPKINS SCHOOL MED.
PI Lee S, McPherson AC;
P-PSDB; R63159.
DR New growth differentiation factor 8 - useful for treatment and
PT diagnosis of cell proliferative disorders esp. of muscle.
PS Claim 3; Page 47; 84pp; English.
CC GDF-8 can be used to maintain cells before transplantation; to
CC improve efficiency of cell fusion and to treat obesity or diseases
CC related to abnormal adipocyte proliferation.
SQ Sequence 2676 BP; 881 A; 506 C; 508 G; 781 T;

Query Match 13.2% Score 23; DB 13; Length 2676;
Best Local Similarity 70.2% Pred. No. 1.76e+00;
Matches 40; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Db 2534 aatacatcttcttacttactgaattataaattggaacttggtatcaatgt 2590
CP 163 AATTCACCTGTTTATTATTATTATCTGCATATAATTCATATTATTATGCAATGT 107

RESULT 11
ID T14322 standard; DNA; 67 BP.
AC T14322;
DE 16-JAN-1997 (first entry)
DE Primer used in the labelling and sorting of nucleotide molecules.
KW Labelling; sorting; sequencing; tag; tagging; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 67
FT /tag= a
FT /mod_base= linked to biotin
PN WO9612039-A1.
PD 25-APR-1996.
PE 12-OCT-1995; U12678.
PR 13-OCT-1994; US-322348.
PR 19-DEC-1994; US-359295.
PA (LYNX-) LYNX THERAPEUTICS INC.
PI Brenner S;
P-PSDB; 96-222023/22.
DR Labelling and sorting mols. using oligo:nucleotide tags - useful in
PT large-scale parallel operations, e.g. DNA sequencing and mRNA
PT fingerprinting
PS Disclosure: Page 19; 71pp; English.
CC Determining the nucleotide sequence (I) of a target polynucleotide
CC (T) comprises: (a) generating from T a plurality of fragments that
CC (T) cover T; (b) attaching an oligonucleotide tag from a repertoire of
CC tags, to each fragment such that all the same fragments have the
CC same tag, and all different fragments have different tags; (c)
CC sorting the fragments by specifically hybridizing the tags with
CC their respective tag complements; (d) determining (I) of T by collating
CC of each of the fragments; and (e) determining (I) of T by collating
CC the sequences of the fragments. The tagging system can be used with
CC single base sequencing methods to sequence polynucleotides up to

```

Query Match	12.68;	Score 22;	DB 34;	Length 68;
Best Local Similarity	13.98;	Pred. No.	4.48e+00;	

KM Antiviral; cell growth regulator; immune system regulator;
DE IFNX 444.
Sequence encoding new modified human interleukin polypeptides

```

KW antiproliferative; ss.
OS Homo sapiens. Location/Qualifiers
FH Key 1..501
FT cds /*tag= a
EP-163993-A.
PN 11-DEC-1985.
PD 17-MAY-1985: 105750.
PE 17-MAY-1984: GB-012564.
PA (SEAR ) SEARLE G D & CO.
PI Bell LD, Boseley PG, Porter AG;
WPI: 85-311944/50.
DR P-PSDB: P50026.
PT New modified human beta interferon polypeptide(s) - prepd. by
PT anti-proliferative and immune regulating actions
PS Claim 28: Chart 2e, page 36; 71pp: English.
CC Compared with interferon beta prepd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index: improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixts.
SQ Sequence 501 BP; 112 A; 31 C; 67 G; 80 T;

Query Match 12.6%; Score 22; DB 3; Length 501;
Best Local Similarity 25.0%; Pred. No. 4.48e+00;
Matches 20; Conservative 28; Mismatches 32; Indels 0; Gaps 0;

Db 12 yytbytgntlyybcarmgdmnmnaaytlycartgycaraarybytbgtgcaryt 71
Cp 153 TTTTATTTTATCTGGCATATTAATTCATATTTTATGTAAGTGGTGTATACATT 94
Cp 93 AGAATAATAGATTACACATT 74

Db 72 baaygsmgdytbgartayt 91
Cp 93 AGAATAATAGATTACACATT 74

RESULT 15
ID N50031 standard; DNA; 501 BP.
AC N50031:
DT 04-SEP-1991 (first entry)
DE Sequence encoding new modified human beta interferon polypeptides
DE IFN $\gamma$  448.
KW Antiviral; cell growth regulator; immune system regulator;
KW antiproliferative; ss.
OS Homo sapiens.
FH Key 1..501 Location/Qualifiers
FT cds /*tag= a
EP-163993-A.
PN 11-DEC-1985.
PD 17-MAY-1985: 105750.
PE 17-MAY-1984: GB-012564.
PA (SEAR ) SEARLE G D & CO.
PI Bell LD, Boseley PG, Porter AG;
WPI: 85-311944/50.
DR P-PSDB: P50030.
PT New modified human beta interferon polypeptide(s) - prepd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PS Claim 28: Chart 2i, page 40; 71pp: English.
CC Compared with interferon beta prepd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index: improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixts.
SQ Sequence 501 BP; 110 A; 30 C; 69 G; 80 T;

Query Match 12.6%; Score 22; DB 3; Length 501;
Best Local Similarity 25.0%; Pred. No. 4.48e+00;
Matches 20; Conservative 28; Mismatches 32; Indels 0; Gaps 0;

```

```

Db 12 yytbytgntlyybcarmgdmnmnaaytlycartgycaraarybytbgtgcaryt 71
Cp 153 TTTTATTTTATCTGGCATATTAATTCATATTTTATGTAAGTGGTGTATACATT 94
Db 72 baaygsmgdytbgartayt 91
Cp 93 AGAATAATAGATTACACATT 74

```

Search completed: Tue Dec 22 04:48:25 1998
Job time : 46 secs.

MUSCULUS DOMESTICUS (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_nm n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Tue Dec 22 04:48:44 1998; Msrch time 235.33 Seconds
1198.708 Million cell updates/sec
Tabular output not generated.

Title: >US-08-530-112A-1320
Description: (1-128) from US08530112A.seq
Perfect Score: 128
N.A. Sequence: 1 GATCAGAAATTTTAAAGCAG.....AGAAATCTTGCTGGGCANA 128
Comp: CTAGCTTTAAAAATTTCTGTC.....TCTATAGACGACCCGTTT

Scoring table:
TABLE default
Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 567134 seqs, 1101898692 bases x 2

Post-processing: Minimum Match 08
Listing first 45 summaries

Database:

emb155
1:em_ba 2:em_fun 3:em_htg 4:em_hun1 5:em_hun2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_vl

Database:

genbank107
15:gb_ba1 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_om 20:gb_ov
21:gb_pat 22:gb_ph 23:gb_pl 24:gb_pl 25:gb_pri
26:gb_pri 27:gb_pri 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
32:gb_un 33:gb_vl

Statistics: Mean 9.131; Variance 5.660; scale 1.613

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	30	23.4	7218	21	166494	Sequence 14 from paten	4.80e-03
2	30	23.4	10772	18	AF012089	Drosophila melanogaste	4.80e-03
3	29	22.7	10772	18	AF012089	Drosophila melanogaste	1.53e-02
4	29	22.7	92879	17	AC004450	*** SEQUENCING IN PROG	1.53e-02
5	27	21.1	2340	24	AF017990	Schistosaccharomyces po	1.49e-01
6	27	21.1	97614	17	AC004937	*** SEQUENCING IN PROG	1.49e-01
7	26	20.3	96801	23	ATAC003033	Arabidopsis thaliana c	4.52e-01
8	25	19.5	747	26	AG000833	Homo sapiens genomic D	1.35e+00
9	25	19.5	100000	26	AP000054	Homo sapiens genomic D	1.35e+00
10	25	19.5	135289	17	AC004961	*** SEQUENCING IN PROG	1.35e+00
11	25	19.5	144518	17	AC004914	*** SEQUENCING IN PROG	1.35e+00
12	25	19.5	216021	26	HUAC004787	Homo sapiens Chromosom	1.35e+00
13	24	18.8	4080	23	SCYL108C	S.cerevisiae Chromosom	3.94e+00

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	BASE COUNT	ORIGIN
1	166494	Sequence 14 from patent US 5670367.	166494	92724471	Unknown.	Unknown.	Unclassified.	1 (bases 1 to 7218)	Donner, F., Scheiflinger, F. and Falkner, F. Gunter.	Recombinant fowlpox virus	Patent: US 5670367-A 14 23-SEP-1997;	Location/Qualifiers	1. 7218	1944 a 1491 c 1486 g 1929 t 368 others

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	BASE COUNT	ORIGIN
2	AF012089	Drosophila melanogaster cysteine proteinase-1 (CP1) gene, complete cds, and phenylalanine tRNA synthetase gene, partial cds.	AF012089	92305220	Unknown.	Unknown.	Unknown.	1 (bases 1 to 7218)	Donner, F., Scheiflinger, F. and Falkner, F. Gunter.	Recombinant fowlpox virus	Patent: US 5670367-A 14 23-SEP-1997;	Location/Qualifiers	1. 7218	1944 a 1491 c 1486 g 1929 t 368 others

Query Match 23.4%; Score 30; DB 21; Length 7218;
Best local Similarity 3.3%; Pred. No. 4.80e-03;
Matches 2; Conservative 43; Mismatches 15; Indels 0; Gaps 0;

AUTHORS	Rounsley,S.D. and Lin,X.
TITLE	Direct Submission
JOURNAL	Submitted (20-MAR-1998) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, rounsley@igr.org On Sep 25, 1998 this sequence version replaced g1:3643587.
COMMENT	*** *** WARNING: Phase 2 High Throughput Genome Sequence *** ***
FEATURES	* This sequence is unfinished. It consists of 1 contigs for * which the order is known. The lengths of the gaps have been * estimated by the submitter but are not known exactly. When * sequencing is complete, the sequence data presented in this * record will be replaced by a single finished sequence * with the same accession number. 1 92879: config of 92879 bp in length. Location/Qualifiers 1..92879 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /clone="F14B2"
BASE COUNT	29450 a 17049 c 16684 g 29624 t 72 others
ORIGIN	
Query Match	22.7%; Score 29; DB 17; Length 92879;
Best Local Similarity	15.6%; Pred. No. 1.53e-02;
Matches 12; Conservative 41; Mismatches 24; Indels 0; Gaps 0;	
Dd 92792	RAWGSGRAWKSSSCYKRKSSCGKWMTTYYCYCTWMSMWYKKRKSCKWMTTYMMM 92851
Oy 14	AAAGGAGATTTTGCAAAAGGTCAATTTTTTGTCITACCCCTTTTACACTTTTCAGATT 73
Dd 92852	CSSMMWWKGKRSNMVCT 92868
Oy	I::: :I: ::::
Oy	CTCAAAGTGTCTCATCT 90

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 92879)

```

AUTHORS      Rounsley,S.D. and Lin,X.
TITLE        Direct Submission
JOURNAL      Submitted (20-MAR-1998) The Institute for Genomic Research, 9712
              Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
              On Sep 25, 1998 this sequence version replaced g1:3643587.
COMMENT      ***
              *** WARNING: Phase 2 High Throughput Genome Sequence ***
              ***
              This sequence is unfinished. It consists of 1 contigs for
              * which the order is known. The lengths of the gaps have been
              * estimated by the submitter but are not known exactly. When
              * sequencing is complete, the sequence data presented in this
              * record will be replaced by a single finished sequence
              * with the same accession number.
              1 92879: contig of 92879 bp in length.
              Location/Qualifiers
                1..92879
                  /organism="Arabidopsis thaliana"
                  /db_xref="taxon:3702"
                  /clone="F14B2"
FEATURES
  source
    BASE COUNT      29450 a 17049 c 16684 g 29624 t      72 others
    ORIGIN
      Query Match      22.7% Score 29; DB 17; Length 92879;
      Best Local Similarity 15.6%; Pred.No.1.53e-02;
      Matches 12; Conservative 41; Mismatches 24; Indels 0; Gaps 0;
Db 92792 RAWGSGRAWGSSSSCYKRWKMTTYYCYTWSSMWYKKSSGKMTTYMMM 92851
Oy 14 AAAGCAGATTTTGCAAAAGCGTCATTTTGTGTCTACCCCTTTACACTTTTCAGATT 73
Db 92852 CSSMMWKKGKRSMMYCT 92868
Oy 74 CTCAAAGTGTCTCATCT 90
RESULT 5
LOCUS      AF017990 2340 bp DNA PLN 10-AUG-1998
DEFINITION Schizosaccharomyces pombe FRBP-type peptidyl prolyl cis-trans
ACCESSION  AF017990
NID        g3406741
KEYWORDS   fission yeast.
SOURCE     Schizosaccharomyces pombe
ORGANISM   Eukaryotae; Fungi; Ascomycota; Archaeascomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE  1 (bases 1 to 2340)
            Himukai,R., Kuzuhara,T. and Horikoshi,M.
            Structural Classification of FRBP-type Peptidyl Prolyl cis-trans
            isomerases(PP1ase)
            Unpublished
            2 (bases 1 to 2340)
            Himukai,R., Kuzuhara,T. and Horikoshi,M.
            Direct Submission
            Submitted (09-AUG-1997) Laboratory of Developmental Biology,
            Institute of Molecular and Cellular Biosciences, The University of
            Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan
FEATURES
  source
    /organism="Schizosaccharomyces pombe"
    /strain="JY741"
    /db_xref="taxon:4896"
    601..1686
    /gene="fkbp39+"
    /gene="fkbp39+"
    /note="FRBP-type peptidyl prolyl cis-trans isomerase"
    /codon_start=1
    /product="Fkbp39p"
    /db_xref="PID:g3406742"
    /translation="MSPIPIAVYISLVKGDPAVEESTDASIHLTMASIDAGEKSNRP

```

```

TTTIVKVRPIVPEDEDEDEFLIDROMELLSEORFVICTIKREGSLYOQBLNLTITPC
DEVFPSASGATTHLSGNFLVDEDEDEDESESDYDUSPIEBEDLVETVSQDESEES
ESDENSASEEDLSDSPAKKQYKRRKESDEEASPKNNTRKQKGVPEKKE
KVAFAEKLEGGPTAPAKKEQOASNSAPKRTILKGCVVTDVDTGSGASATNGK
KVMRYIGKLENGKRVFNTKGFPAFILLRGEVIRGMDVAGMGGGERKITIPAP
MAYGNOSIPGIPNSITLVEFKLVRYH"

BASE COUNT      766 a      396 c      462 g      716 t
ORIGIN

Query Match      21.1%; Score 27; DB 24; Length 2340;
Best Local Similarity 73.7%; Pred. No. 1.49e-01;
Matches 42; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 1428 AAAAGTTGAATGAGATATATGGAAGCTCGAAATGGAAGTTTGTGCAAAAA 1484
CP 98 AAAAGTTGAGATGAGACACTTGAGAACTGAAAAAGTGTAAAGGGGAGACAAAA 42

RESULT 6 AC004937 97614 bp DNA HTG 12-JUN-1998
DEFINITION *** SEQUENCING IN PROGRESS *** Homo sapiens clone DU0960118; HTGS
ACCESSION AC004937
NID 93213060
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS 1 (bases 1 to 97614)
TITLE The sequence of Homo sapiens clone
JOURNAL 2 (bases 1 to 97614)
AUTHORS Waterston, R.H.
REFERENCE Submitted (12-JUN-1998) Genome Sequencing Center, Washington
TITLE Direct Submission
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

*** WARNING: Phase 1 High Throughput Genome Sequence ***
***
*** This sequence is unfinished. It consists of 5 contigs for
*** which the order is not known; their order in this record is
*** arbitrary. In some cases, the exact lengths of the gaps
*** between the contigs are also unknown; these gaps are presented
*** as runs of N as a convenience only. When sequencing is complete,
*** the sequence data presented in this record will be replaced
*** by a single finished sequence with the same accession number.
***
2446 2445: contig of 2445 bp in length
2464 2463: gap of unknown length
2464 12423: contig of 9960 bp in length
12424 12441: gap of unknown length
24719 24719: contig of 12278 bp in length
24737 24737: gap of unknown length
24738 44189: contig of 19452 bp in length
44190 44207: gap of unknown length
44208 97614: contig of 53407 bp in length.
Location/Qualifiers
1. 97614
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DU0960118"

BASE COUNT      27971 a      20178 c      20167 g      29226 t      72 others
ORIGIN

Query Match      21.1%; Score 27; DB 17; Length 97614;
Best Local Similarity 70.1%; Pred. No. 1.49e-01;
Matches 47; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db 4349 GTCATCCACTTTTACCTGACATCTCATATTCATTCATCTTTAGCAAAAT 4408
OY 47 GTCATCCCTTTTACATCTTTCAGATTCCTCAAAAGTCTCTCATCTCAACTTTTAAAGAAAT 106

```

```

Db - 4409 AAATPAT 4415
OY 107 AAAGAAAT 113

RESULT 7
LOCUS ATAC003033 96801 bp DNA PLN 19-DEC-1997
DEFINITION Arabidopsis thaliana chromosome II BAC T21L14 genomic sequence,
ACCESSION AC003033
NID 92702261
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
AUTHORS Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
Mangoliopsida; Caprales; Brassicaceae; Arabidopsis.
1 (bases 1 to 96801)
Rounsley, S.D., Lin, X., Ketchum, K.A., Crosby, M.L., Brandon, R.C.,
Sykes, S.M., Kaul, S., Mason, T.M., Kerlavage, A.R., Adams, M.D.,
Somerville, C.R. and Venter, J.C.
Arabidopsis thaliana chromosome II BAC T21L14 genomic sequence
Unpublished
2 (bases 1 to 96801)
Rounsley, S.D. and Lin, X.
Direct Submission
Submitted (29-OCT-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
3 (bases 1 to 96801)
Rounsley, S.D.
Direct Submission
Submitted (19-DEC-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Dec 19, 1997 this sequence version replaced g1:2661151.
Address all correspondence to:
Steve Rounsley
The Institute for Genomic Research
9712 Medical Center Dr.,
Rockville, MD 20850,
USA
e-mail: rounsley@tigr.org
BAC clone T21L14 is from Arabidopsis chromosome II and contains the
molecular marker COPI.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL (available by anonymous ftp
from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://genomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
(http://www.cbs.dtu.dk/netgene/cbsnetgene.html), searches of the
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/tdb/at.html).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
genomic sequence that are not annotated as genes but have predicted
exons by GRAIL are annotated as misc features.
Location/Qualifiers
1. 96801
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"

```



```

misc_feature
/clone="T21L14"
/complement(<1..>1616)
/gene="T21L14.1"
/misc_feature
/clone="T21L14.1"
/complement(<1..>1616)
/gene="T21L14.1"
/gene="T21L14.1"
/complement(<1..>1616)
/gene="T21L14.1"
/codon_start=1
/product="putative disease resistance protein (cf-2.2), 3'
partial"
/db_xref="PID:92702262"
/translation="MRHFCSLLLLYCIVFVSSFLTTDLACLPDIOIALIOFKNEFE
SDGCRSDIYNOCDDNTGATKLOPSCCTGKIKPSSSLPELQARLYNTSINN
TSSSPSEFNLTRELVSLASSSTGCVPSISMLITLHNLNLSHNEIGTFPVKN
LTKSLFDLSTNDFSGALPDLPLPLSLYDLKKNHLSIDVNSSSSKLVRLS
LGFNOFEKRIIEPIKSLINLNLHLEASINSHPIDLRVAPLAKSLVDFIDRONRLPA
LSLSDSEFPLSLILIQCDIEEPNIFKTONLEHIDISNNLIKGVPEWFKLPR
LSIANLVNNSLTGFESEVLNLSVOLDFVNGMTGAFPPPLGSISYLSAMNSFT
GNIPLSICNBSLIYDLISYKFTGPICOLSLNKLKVNLRKNSLGSIPDEHSCAKT
OTLDVGIRLTKLPEKSLNCSLNFSLVDNRRIDTFPWLKALPNLHLVTLRSNR
FGLSPDPGRPLAEPFLRIELSDNSFTGS"
1..16183
/misc_feature
/clone="overlapp with BAC clone F25118
(AC002334:76975..93158)."
2100..2205
/note="exon predicted by xgrail, quality
excellent_shadowexon"
join(2339..2536,2938..2981,3093..3245,3444..3557,
3640..3767,3879..4001,4101..4220,4310..4714)
/gene="T21L14.2"
2339..4714
/note="identical to GB:D88374"
/gene="T21L14.2"
join(2445..2536,2938..2981,3093..3245,3444..3557,
3640..3767,3879..4001,4101..4220,4310..4507)
/gene="T21L14.2"
/codon_start=1
/product="mitochondrial F1-ATPase, gamma subunit"
/db_xref="PID:92702263"
/translation="MAVFRREGRLLPSTIAARPIAIAIRSPSSDOEGILLGVRSISTQ
VVRNRMKSVKNIOKITKAMKMAAASKLRAVOGRANSGLMOPFTALLGDNSISYVK
SVVYLTSDKGLCGINSTVYVSRALYKLNAPKEKVOFVYGEKAKARIMRSDKN
IVLSTELKNKPNLAYOVSYLADDLIKNVEPALIYVNRKFSVAFPTYSTVTSPE
IIEKSEIGKIGELDYIEIGETKGEILQWLAFOFSCVFENAVLNNACSEMGARM
SAMSSSRNAEMEDRLTLTYNRQASITTELIBIISGASALEAK"
<5145..>5801
/gene="T21L14.3"
<5145..>5801
/gene="T21L14.3"
5145..5801
/gene="T21L14.3"
/codon_start=1
/product="similar to disease resistance protein (cf-2.2)"
/db_xref="PID:92702264"
/translation="MITYKNAGSISYQDFIDRYGLHMEGRITLVSADPSC
NRBQIPESIGLKALIALNLSNNAFTGNITPMANALIESLDSNNGSLGTPPG
LKTLSFLGYINVSIMHOLKEIPQGIQPPSSPEGNAGLGLPLESCFCTKVPPI
QOSKEDNQEDAKVLNMRKAVATGYGPFGLAIAQIIASYPEWLKIGPNKRNRH"
misc_feature
/complement(5845..5920)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
5905..6010
/note="exon predicted by xgrail, quality good"
/complement(6037..6128)
/note="exon predicted by xgrail, quality good_shadowexon"
join(<6869..7386,7435..8328,8467..>9611)
/gene="T21L14.4"
<6869..>9611
/gene="T21L14.4"
join(6969..7386,7435..8328,8467..9611)
/gene="T21L14.4"

```

```

misc_feature
/clone="T21L14.5"
/complement(<10239..>12593)
/note="predicted by genefinder"
/gene="T21L14.5"
/complement(join(10239..10625,10761..10827,10934..11475,
11535..11763,11869..12593))
/note="hypothetical protein"
/gene="T21L14.5"
/codon_start=1
/db_xref="PID:92702266"
/translation="MTNMEYDRISDENGATREALTYLEAYDNLALAMEACGKTL
PSVAKSSPVSVSTBOSTAEBSATPLTIPRVSNQDPSINTPLSTLNPAPKRS
HSGRVLPPOELSHLQERINFLAVASDLPSOELVLEFEMNDVRAQAQVYEDR
DOORTSYDREIDOSYLQFOEGYPMGVNLEGSQNETMDLMOHSAPLSLAL
LPESOFHDSVGYOLOSSEFSGMLANDVDEFGVYDIPIMSSSSGOVYLIFPSAL
YLSGLCELFKIVFKKGTNMLIDAHNSFYVDEGCTEKTIRIOSSSSAQNEDHP
PNNDRICTERGVSNGEAVYILESDGDLAIAIEKRSKNPTLAVRSSPVYSVEIS
APWESPTERPLVNNSCPISNELOIOEQLPNEELIKQCELVGLHPDVAAYLE
RCQWLSQVAIDYFENAEYFEDVYSNPPLIGLPSQOFQASGSSFSLNINPDTF
TVANLPNASSOVDOPRYISLDGLSLDVAFLFVEEDLSTETFDPAIOGWSVGHIEA
PTTITLYLDDGGSVTPEYIPERSDQVDIRNAIDELTPANNRYDLRSVSGEDDC
RDMNATVCKIKSGSTLHQVFHN"
13033..13099
/repeat_region
/rpt_family="AT_rich"
/complement(13243..13275)
/rpt_family="AT_rich"
13384..13468
/note="exon predicted by xgrail, quality good_shadowexon"
13348..13682
/marginal_shadowexon
13908..14048
/note="exon predicted by xgrail, quality
excellent"
misc_feature
/complement(14060..14217)
/rpt_family="AT_rich"
/complement(14291..14336)
/rpt_family="AT_rich"
14509..14567
/repeat_region
Note: remainder of annotations omitted.
Query Match 20.3% Score 26; DB 23; Length 96801;
Best Local Similarity 69.18; Pred. No. 4,52e-01;
Matches 47; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
Db 17542 AAAAGTTGAGAGAAAACCTACATATAGAAATGCAAAATGCGTCAAAATTAAC 17601
|||||
Cp 98 AAAAGTTGAGATGACATTTGAGATCTGAAAAGGTGACACAAAAAAT 39
Db 17602 CACATTTT 17609

```

```

CP      38 GACCCTTT 31
|||||
RESULT 8
LOCUS AG000833 747 bp DNA PRI 31-OCT-1997
DEFINITION Homo sapiens genomic DNA, 21q region, clone: PQ31L12B671.
ACCESSION AG000833
NID 92579641
KEYWORDS
SOURCE Homo sapiens DNA, clone: PQ31L12B671.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Hattori,M., Ishii,K., Shiba,T. and Sakaki,Y.
TITLE Homo sapiens genomic DNA, chromosome 21q
JOURNAL Published Only in Database (1997) In press
REFERENCE
AUTHORS Hattori,M., Ishii,K., Shiba,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-1997) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, Kitasato University, Department of Science, JST
Sequencing Laboratory, Kitasato 1-15-1, Sagamihara 228, Japan
(E-mail:hattori@ngc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,
Fax:0427-78-9732)
FEATURES
source
Location/Qualifiers
1..747
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="PQ31L12B671"
/map="21q"

BASE COUNT 210 a 164 c 155 g 204 t 14 others
ORIGIN
Query Match 19.5%; Score 25; DB 26; Length 747;
Best Local Similarity 71.9%; Pred. No. 1.35e+00;
Matches 41; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 80 ATCTTTATTTGTTAGATTATGACATGAGATTATTCCTCATCTGTAATAATGTAAA 136
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CP 113 ATCTTTATTTGTTTAAAGTTGAGATGAGACACTTGAGATCTGAGAAAGTGTTAAA 57

RESULT 9
LOCUS AP000054 100000 bp DNA PRI 13-MAY-1998
DEFINITION Homo sapiens genomic DNA, chromosome 21q11.1, segment 25/28,
complete sequence.
ACCESSION AP000054
NID 93132364
KEYWORDS
SOURCE HTG.
ORGANISM Homo sapiens DNA, clone:245P17-f4A4f_4.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
TITLE Homo sapiens genomic DNA, chromosome 21q
JOURNAL Published Only in Database (1998) In press
REFERENCE
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (11-MAY-1998) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, Kitasato University, Department of Science, JST
Sequencing Laboratory, Kitasato 1-15-1, Sagamihara 228, Japan
(E-mail:hattori@ngc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,
Fax:0427-78-9561)
FEATURES
source
Location/Qualifiers
1..100000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"

CP      38 GACCCTTT 31
|||||
BASE COUNT 28599 a 21687 c 22052 g 27662 t
ORIGIN
Query Match 19.5%; Score 25; DB 26; Length 100000;
Best Local Similarity 71.9%; Pred. No. 1.35e+00;
Matches 41; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 79524 TTTCACATTTTACAGATGAGGAATATCATCTGCAATCAATCAATCAATCAAT 79580
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CP 57 TTTCACATTTTACAGATTTCTCAAAAGTGTCTCATCTTCACTTTAAAGAAATGAAT 113

RESULT 10
LOCUS AC004961 135289 bp DNA HTG 12-JUN-1998
DEFINITION *** SEQUENCING IN PROGRESS *** Homo sapiens clone DJ0886008; HTGS
ACCESSION AC004961
NID 93213083
KEYWORDS
SOURCE HTG; PHASE1.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS Waterston,R.H.
TITLE 2 (bases 1 to 135289)
JOURNAL Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis
MO 63108, USA
COMMENT
***
*** WARNING: Phase 1 High Throughput Genome Sequence ***
***
* This sequence is unfinished. It consists of 2 contigs for
* which the order is not known; their order in this record is
* arbitrary. In some cases, the exact lengths of the gaps
* between the contigs are also unknown; these gaps are presented
* as runs of N as a convenience only. When sequencing is complete
* the sequence data presented in this record will be replaced
* by a single finished sequence with the same accession number.
* 1 10000: contig of 10000 bp in length
* 10001 10018: gap of unknown length
* 10019 135289: contig of 125271 bp in length.
FEATURES
source
Location/Qualifiers
1..135289
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DJ1098J04"

BASE COUNT 37377 a 32196 c 30320 g 35378 t 18 others
ORIGIN
Query Match 19.5%; Score 25; DB 17; Length 135289;
Best Local Similarity 71.9%; Pred. No. 1.35e+00;
Matches 41; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 3208 AAGTTTAAAGTAGAAGCAAAATTTGACATAGTCAATATGATCTTTGAAA 3264
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CP 65 AAGTGTAAAGGGGTAGACAAAATAAGACCCCTTTGACAAAATTCGCTTTAAAA 9

RESULT 11
LOCUS AC004914 144518 bp DNA HTG 12-JUN-1998
DEFINITION *** SEQUENCING IN PROGRESS *** Homo sapiens clone DJ0886008; HTGS
ACCESSION AC004914
NID 93213083
KEYWORDS
SOURCE HTG; PHASE1.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS Waterston,R.H.
TITLE 2 (bases 1 to 135289)
JOURNAL Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis
MO 63108, USA
COMMENT
***
*** WARNING: Phase 1 High Throughput Genome Sequence ***
***
* This sequence is unfinished. It consists of 2 contigs for
* which the order is not known; their order in this record is
* arbitrary. In some cases, the exact lengths of the gaps
* between the contigs are also unknown; these gaps are presented
* as runs of N as a convenience only. When sequencing is complete
* the sequence data presented in this record will be replaced
* by a single finished sequence with the same accession number.
* 1 10000: contig of 10000 bp in length
* 10001 10018: gap of unknown length
* 10019 135289: contig of 125271 bp in length.
FEATURES
source
Location/Qualifiers
1..135289
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DJ1098J04"

BASE COUNT 37377 a 32196 c 30320 g 35378 t 18 others
ORIGIN
Query Match 19.5%; Score 25; DB 17; Length 135289;
Best Local Similarity 71.9%; Pred. No. 1.35e+00;
Matches 41; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 3208 AAGTTTAAAGTAGAAGCAAAATTTGACATAGTCAATATGATCTTTGAAA 3264
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CP 65 AAGTGTAAAGGGGTAGACAAAATAAGACCCCTTTGACAAAATTCGCTTTAAAA 9

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 144518)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 144518)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

*** WARNING: Phase 1 High Throughput Genome Sequence ***

* This sequence is unfinished. It consists of 1 contigs for
* which the order is not known; their order in this record is
* arbitrary. In some cases, the exact lengths of the gaps
* between the contigs are also unknown; these gaps are presented
* as runs of N as a convenience only. When sequencing is complete,
* the sequence data presented in this record will be replaced
* by a single finished sequence with the same accession number.
1 144518: contig of 144518 bp in length.
Location/Qualifiers
1..144518
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="D00886008"
BASE COUNT 39428 a 31608 c 31519 g 41963 t
ORIGIN

Query Match 19.5%; Score 25; DB 17; Length 144518;
Best Local Similarity 71.9%; Pred. No. 1.35e+00;
Matches 41; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 90116 AGCTTAAAGTAGACGCAAAATGACATCTGATCTTGA 90172
CP 65 AAGTGTAAAGGGGTACACAAATGACCTTTGACAAATTCCTTTAAAA 9

RESULT 12
LOCUS HUAC004787.216021 bp DNA PRI 24-JUL-1998
DEFINITION Homo sapiens chromosome 16 BAC clone CIT987SK-A-952F10, complete
sequence.
ACCESSION AC004787
NID 93337381
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 216021)
AUTHORS Adams, M.D., Loftus, B.J., Zhou, L., Crosby, M., Fuhrmann, J.,
Mason, T.M., Brandon, R., Kim, U.J., Kerlavage, A.R. and Venter, J.C.
TITLE Homo sapiens chromosome 16 BAC clone CIT987SK-A-952F10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 216021)
AUTHORS Adams, M.D. and Loftus, B.J.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, Email:
bjloftus@tigr.org
3 (bases 1 to 216021)
AUTHORS Adams, M.D. and Loftus, B.J.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Jul 24, 1998 this sequence version replaced g1:3241936.
Address all correspondence to: Mark Adams The Institute for Genomic
Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
address: huntgen@tigr.org. The orientation of the sequence is from
sp6 end to 7 end. Genes were identified by a combination of five

methods including: XGRail (available by anonymous ftp from
artthur.gem.ornl.gov), GeneFinder (Phil Green, University of
Washington), GenScan (Chris Burge,
http://genomic.stanford.edu/~chris/GENSCAN.html) searches of the
complete sequence against a peptide database, and the Human gene
index database at TIGR (http://www.tigr.org/tdb/iggi.html).
Genes without peptide homology having spliced EST hits are termed
'unknown gene product'. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Location/Qualifiers
1..216021
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q21.22"
/clone="A-952F10"
27765..27872
/note="7766, STS1-CSRL-27g3-uA/CSRL-27g3-uZ, Chr. -, Homo
sapiens"
/db_xref="dbSTS:G02280"
73826..73943
/note="7608, STS1-CSRL-24g1-uA/CSRL-24g1-uZ, Chr. -, Homo
sapiens"
/db_xref="dbSTS:G02122"
175801..175945
/note="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"
/db_xref="dbSTS:G09703"
175810..175945
/note="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"
/db_xref="dbSTS:G09935"
199463..199572
/note="9824, WI-3555, Chr. 16, Homo sapiens"
/db_xref="dbSTS:G04338"
BASE COUNT 60960 a 51778 c 49172 g 53987 t 124 others
ORIGIN

Query Match 19.5%; Score 25; DB 26; Length 216021;
Best Local Similarity 5.4%; Pred. No. 1.35e+00;
Matches 2; Conservative 29; Mismatches 6; Indels 0; Gaps 0;

Db 1457 KKYMKSMRRARRSGAKKKKYYCYCYCYCYCY 1493
CY 23 TTTGTCAAAAGGTCATTTTGTCTACCCCTTTT 59

RESULT 13
LOCUS SCYJ108C.4080 bp DNA PLN 11-AUG-1997
DEFINITION S.cerevisiae chromosome X reading frame ORF YJL108C.
ACCESSION Z49383 Y13136
NID 91008289
KEYWORDS baker's yeast.
SOURCE Saccharomyces cerevisiae
ORGANISM Saccharomyces cerevisiae
Eukaryota; Eukaryota; Eukaryota; Fungi; Ascomycota;
Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;
Saccharomycetes.
REFERENCE 1 (bases 1 to 4080)
AUTHORS Rasmussen, S.W.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4080)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-1995) Data collected by MIPS on behalf of the
European yeast chromosome X sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitze 18a D-82152
Martinsried, FRG; E-mail: Mewesen@mips.emblnet.org
Location/Qualifiers
1..4080
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="X"
complement(396..1547)
/note="ORF YJL108C"

```
/codon_start=1
/db_xref="pid:g1008290"
/db_xref="SWISS-PROT:p42946"
/translation="MVSFGDATTTRTSEVQLVRCQGLNLMKLVHGVHAYKRVHDTL
GADENALIDQIADIDNLVPMWCVLLAFCSAMVTPYAFGADWNLASIFPMGLCVG
SLOFELISOKSYMSNVFEISASIVSFCGRFAGSIIPRHICGATGGLALILGVI
ILCGLEIORSRLVAGVAMFAYLIYSLEIGITLGLSLFGWMTNATNEISCPLI
SPWRFLEVPAPFTISILNOAHISQLPVMTISCTGYGVYTWACKHFNSTFPAAL
AAFYIGVGLNLSRIKMKLAVSAMLPALFYVPPSGIASNSLSLGLQSNNTLVNANET
ITTSIDSPSSSKSEFGMTMIOVCGLISVGLFASLFLVYPGRKKTGLFSL"
complement(1544..2707)
/note="ORF YJL107c"
/codon_start=1
/db_xref="pid:g1008291"
/db_xref="SWISS-PROT:p42947"
/translation="WDGRNEKFTTPVSDFRSGSSSEQAGVNLIEDSSDHRTSASBKS
KGNLSGKSIDGINSNDKNKNTFADIDALENDLSRSTETSDNSKSGIDGDEDRP
ARHKRPVSEFHLNNGKDGDETFIKIINNNGNOGLVPGIAPIPSENGKND
LEKNNRNEIPLSDIADASKIVDHEGDDKELEALKLEGDVNCTSDGTGLSSKNS
FLAPVHDHDDYAKENNSDDNEGFTETSTYVPPSOVKSGVLSLKLITQNDONSSS
IFSDQAVTTDEGISSTAGNCDVPYAKRSIQNKGRKKRMRPLKRLTEAKIT
VHADIIDRHRILMKRCALMRTGAPTHRLERTMWTSTKVLIDGFCIFQVY"

BASE COUNT      1268 a      803 c      767 g      1242 t
ORIGIN

Query Match      18.8% Score 24: DB 23: Length 4080;
Best Local Similarity 69.4%; Pred. No. 3.94e+00;
Matches 43: Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Db 2197 GAATTTCTTATTCAGCGTATTTTTCATATCATTCCTCCATTTCTGTAAG 2256
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 20 GAATTTGCAAAAGGCTATTTTGTCTACCCCTTTACACTTTTCGATTCGAA 79
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 2257 GT 2258
||
Oy 80 GT 81

RESULT 14
LOCUS CEF56A12 26028 bp DNA INV 21-SEP-1998
DEFINITION Caenorhabditis elegans cosmid F56A12, complete sequence.
ACCESSION Z81551
NID g1914348
KEYWORDS HTG; Homeobox protein.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1. (bases 1 to 26028)
AUTHORS Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
Latteille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopar, A.,
Saunders, D., Showkhen, R., Smaildon, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sprat, J. and Wohlman, P.
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 26028)
AUTHORS Mortimore, B.
TITLE Direct Submission
JOURNAL Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jess@sanger.ac.uk or twenematode.wustl.edu
On Mar 30, 1997 this sequence version replaced g1:166602.
Coding sequences below are predicted from computer analysis, using
```

predictions from GeneFinder (P. Green, U. Washington), and other available information. For a graphical representation of this sequence and its analysis see: -
<http://webcye.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence&object=F56A12>
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone F56A12. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone F56A12 is at 1 in this sequence. The true left end of clone C15H11 is at 25925 in this sequence. The end of this sequence (23925..26028) overlaps with the start of sequence Z81035.

```
FEATURES
source
Location/Qualifiers
1..26028
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="V"
/clone="F56A12"
/gene="ceh-35"
join(14802..15104,15673..15802,16175..16563,16988..17212)
/gene="ceh-35"
/feature="Similarity to Mouse SIX3 protein (TR:062233)"
/feature="start=1"
/db_xref="PID:g325270"
/db_xref="PID:g2226398"
/feature="start=1"
/codon_start=1
/db_xref="PID:g1914349"
/db_xref="PID:g1914348"
/translation="MNDMDNMTFTKLYSCQYRASDPSQPTIMLQDVPVTSRTS
AIGTYAIFLNGHNRMSALITLCAMDMLNTTYLFLSTELSLRTNTEY
OHQDMLILYGRNSFAMSSPMLVCTITIRRVANNPKLISHGRSRLKISASAKM
STAAOPSSSTESAKFTISPAEMEFYEHRTSRNSGKRANRREFRPLVILVLCFV
IHSTLFEFENLITCFEDVETESKMLQMLQDGSYVQFVVALMTMETETLQPMFI
STLSLETKHMONVKKERRRLEPESOKRSNTLVTEELKAKSKALAVFIVVLELIRS
LPTLIDLYEVLTENGINRPPMTKTVRISDFLIINSATNTLAYFGKVAFFECFRMLE
RRIRCRIVKKEKELISTSLTG"

BASE COUNT      8672 a      4604 c      4726 g      8026 t
ORIGIN

Query Match      18.8% Score 24: DB 18: Length 26028;
Best Local Similarity 67.1%; Pred. No. 3.94e+00;
Matches 47: Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Db 8886 AGAATTTGTCACGTTAAAAAGCCTTACTGGAATTCACACTTTGACAAAATTCTT 8945
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

CP 75 AGAATCTGAAGAGGTAAAGGGTAGACAAAAATGACCCTTTTGACAAAATTCTGCT 16
 Db 8946 TCAAGATTTC 8955
 Cp 15 TTAAAAATTTC 6

RESULT 15
 LOCUS CELF26G5 32087 bp DNA INV 26-FEB-1998
 DEFINITION Caenorhabditis elegans cosmid F26G5.
 ACCESSION AF022974
 MID 92384840
 KEYWORDS
 SOURCE
 ORGANISM
 . Caenorhabditis elegans strain-Bristol N2.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;
 Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 32087)
 Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,
 Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
 Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
 Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
 Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
 Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
 O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A.,
 Saunders, D., Showkhen, R., Smaldon, N., Smith, A., Sonhammer, E.,
 Staden, R., Sulston, J., Thierly-Mieg, J., Thomas, K., Vaudin, M.,
 Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
 Wilkinson-Sprat, J. and Wohldman, P.
 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans
 JOURNAL Nature 368 (6466), 32-38 (1994)
 MEDLINE 94150718
 REFERENCE 2 (bases 1 to 32087)
 AUTHORS Sammons, L., Wohldmann, P. and Beck, C.
 TITLE The sequence of C. elegans cosmid F26G5
 JOURNAL Unpublished (1998)
 COMMENT Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RO, England
 e-mail: tw@nematode.wustl.edu and jesus@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate
 chemistry; an attempt was made to resolve all sequencing problems,
 such as compressions and repeats; all regions were covered by
 sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is C02E7, 1200 bp overlap; 3' cosmid is C13D9, 200 bp
 overlap. Actual start of this cosmid is at base position 1197 of
 CELF26G5; actual end is at 10235 of CELF13D9. This cosmid lies in
 an unanchored cluster, the orientation of which is unknown.

NOTES:

Coding sequences below are predicted from computer analysis, using
 the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES
 source
 1..32087
 /organism="Caenorhabditis elegans"

gene
 CDS
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /chromosome="V"
 /clone="F26G5"
 /gene="F26G5.11"
 complement(83..2720)
 /gene="F26G5.11"
 complement(join(83..534,2031..2231,2432..2720))
 /gene="F26G5.11"
 /note="similar to C. elegans olfactory receptor ODR-10
 (MID:g1235900)"
 /codon_start=1
 /evidence=not_experimental
 /db_xref="pid:g2911714"
 /translation="MSWEQVYVLSQSTGLALIHSLNLNIIHNSKRDIGERYL
 MFSISLEIIVAILDGLVQPSKSTKGLLILLETSSLSIVAGTLMIFTVTYTCGF
 YGSSMAILIHFAVRLMIIISGSEAMLKWPRPKVILMLFPGTMIFVFAVEYCY
 PRTHFWLLESEVDCGLSILICYFEGKCKMSLIFSSSTYOKLOLFYAL
 VAOTLIPVLIHPVLTMTLSPLINDIGALSIVSITIALPFTLDPPLMLIVCHYR
 NATNFRSRKXEFKRVKCGKCLRTKRRHPSSITL"
 3356..5445
 /gene="F26G5.5"
 join(3356..3547,4453..4540,4589..4802,4857..4998,
 5047..5445)
 /gene="F26G5.5"
 /note="similar to C. elegans olfactory receptor ODR-10
 (MID:g1235900)"
 /codon_start=1
 /evidence=not_experimental
 /db_xref="pid:g2384844"
 /translation="MSAMONFQKRAQIFGVIFALLHNLILPITEKSHKELGYKNL
 IIVAISECFYAVLETVRPVHSRGFTFALLIWNDSLELGVKLLAMTCAFGS
 FLVIESVQFIFRWAVSGNKHKEFEGFLITVTLPSIFCGILMWLVAAFPCKRST
 DEYLRDSVKRENLSENLVAPYFVNTDGTIDIVYPSFISILITVLVNIYSIT
 VEFYGIKCYSLREGALVSONTQKLOLQFSLVQIVYPLFLMFPAAAMCEFFL
 EIDVGSIGSIVYTLAIFPAIDPLPLIMWKYSRPAIOEYIVLFFETIRKMLQPSQDN
 SIAMTOTRAI"
 6155..7456
 /gene="F26G5.4"
 join(6155..6343,6395..6482,6527..6740,6861..7002,
 7049..7456)
 /gene="F26G5.4"
 /note="similar to C. elegans olfactory receptor ODR-10
 (MID:g1235900)"
 /codon_start=1
 /evidence=not_experimental
 /db_xref="pid:g2384843"
 /translation="MLDMKLFQVYOLIGSVIGLHLSLISLVNKSPOOMRAYKNLI
 ILSLPEINYSVLELIVOPMPSFGSAPAMVINVECYIDRNVIMLVSYFNAPVSL
 LAMFCIQFTRYRIWAIISGNIMLEKFNGLRMALVYLCGALIMVFTYFPCSRPSAD
 EYLKRNIMNVFNLDIDKNLYIAPYFETNSDGTIDYPSFIAIIVNIIITVLSLV
 FYFGFCYISLQKICKONVSQNTQKQOLFSLIQTPLFPLAMIPCCAMLESLFL
 TIDVGSFTGIVYVTLAIFPAVDPLPLIMWKYCRNLSRYLRLVLSAVSKMFTPDVP
 ANTELISKYTRL"
 complement(8204..8753)
 /gene="F26G5.8"
 complement(join(8204..8457,8555..8753))
 /gene="F26G5.8"
 /codon_start=1
 /evidence=not_experimental
 /db_xref="pid:g2384846"
 /translation="MLPYLCITIKKFAKKTFTSYNDLILVINADHQRKCTVFCALIMKIQ
 NLETNVTMLNTELEYRPHDSNNVTLISQPIRSRPNKFLVMEKCVCSIFNMLW
 EMLSTIIFHLTTLVTFEFAHSEPIFFRHSISFLSEVLEMFVGFCLL"
 8882..10311
 /gene="F26G5.3"
 join(8882..8901,9464..9866,10123..10311)
 /gene="F26G5.3"
 /codon_start=1
 /evidence=not_experimental
 /db_xref="pid:g2384845"
 /translation="MHRATWILITDGSSTVITIPGPTLIGPACVAGHMPGSSIEH
 NLIWLDCLFRYIVYVDPSPKSIIRPAVIVYPSFPHAIWKLPDAGOMEIKAN
 GHELLANQLEFSBMLKTRIVYWSISIVEVQLVLTALVILTFQVCIPIMFLGVFFL

 WISE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_un n.a. - n.a. database search, using Smith-Waterman algorithm
 Run on: Tue Dec 22 04:57:19 1998; MasPar time 36.16 Seconds
 Tabular output not generated. 481.639 Million cell updates/sec

Title: >US-08-530-112A-1320
 Description: (1-128) from US08530112A.seq
 Perfect Score: 128
 N.A. Sequence: 1 GATCGAATTTTAAAGCAG.....AGCATATCTTGCTGGCGCAA 128
 Comp: CTAGCTTAAAAAATTTCGTC.....TCTTATAGAACGACCCGGTTF

Scoring table: TABLE default
 Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: n-geneseq32
 1:part1 2:part2 3:part3 4:part4 5:parts 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39 40:part40

Statistics: Mean 7.096; Variance 5.269; scale 1.347

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description	Pred. No.
1	128	100.0	128 19	T20320 Human gene signature	5.46e-56
2	35	27.3	91 9	Q51746 Oligonucleotide probe	2.38e-06
3	35	27.3	204 1	N81164 Base substituted E.co	1.61e-03
4	29	22.7	91 9	Q51746 Oligonucleotide probe	3.70e-02
5	26	20.3	204 1	N81164 Base substituted E.co	2.83e-01
6	24	18.8	114 12	Q70473 Generic DNA sequence	7.67e-01
7	23	18.0	591 3	N50025 Sequence encoding new	7.67e-01
8	23	18.0	1697 18	T11677 DNA encoding a protei	7.67e-01
9	23	18.0	1953 6	Q36654 Bt isolate 43F coding	7.67e-01
10	23	18.0	1953 24	T43223 Antiscarab pest toxin	7.67e-01
11	23	18.0	2190 4	Q27945 Sequence of CytIIIC(b	7.67e-01
12	23	18.0	2425 1	O05679 btPGS1208 gene.	7.67e-01
13	22	17.2	74 21	T13613 DC43 TSAR library gen	2.05e+00

c	14	22	17.2	81 21	T13611 DC43 TSAR library gen	2.05e+00
c	15	22	17.2	114 12	Q70468 Generic DNA sequence	2.05e+00
c	16	22	17.2	114 12	Q70472 Generic DNA sequence	2.05e+00
c	17	22	17.2	114 12	Q70465 Generic DNA sequence	2.05e+00
c	18	22	17.2	114 12	Q70470 Generic DNA sequence	2.05e+00
c	19	22	17.2	1087 23	T27350 Meloidogyne resistant	2.05e+00
c	20	22	17.2	1103 23	T27351 Meloidogyne sensitive	2.05e+00
c	21	21	16.4	498 3	N50034 Sequence encoding new	5.38e+00
c	22	21	16.4	501 3	N50031 Sequence encoding new	5.38e+00
c	23	21	16.4	501 3	N50030 Sequence encoding new	5.38e+00
c	24	21	16.4	501 3	N50029 Sequence encoding new	5.38e+00
c	25	21	16.4	501 3	N50024 Sequence encoding new	5.38e+00
c	26	21	16.4	501 3	N50022 Sequence encoding new	5.38e+00
c	27	21	16.4	501 3	N50026 Sequence encoding new	5.38e+00
c	28	21	16.4	501 3	N50028 Sequence encoding new	5.38e+00
c	29	21	16.4	1146 40	V24778 H. pylori ORF 049p112	5.38e+00
c	30	21	16.4	1977 1	N91495 Bacillus thuringiensis	5.38e+00
c	31	21	16.4	5125 5	Q29703 IRS-1	5.38e+00
c	32	21	16.4	5775 40	V20437 Human c-K-ras oncogen	5.38e+00
c	33	20	15.6	114 12	Q70469 Generic DNA sequence	1.39e+01
c	34	20	15.6	114 12	Q70467 Generic DNA sequence	1.39e+01
c	35	20	15.6	501 3	N50023 Sequence encoding new	1.39e+01
c	36	20	15.6	565 6	Q35072 HCV envelope region n	1.39e+01
c	37	20	15.6	1047 13	Q78152 Tubex specific class	1.39e+01
c	38	20	15.6	1289 16	Q87016 Clone B2 encodes func	1.39e+01
c	39	20	15.6	1975 33	T51739 Orange pectin methyl	1.39e+01
c	40	20	15.6	2345 17	Q88167 Nr (never ripe) ethyl	1.39e+01
c	41	20	15.6	2404 17	Q88167 Tomato ethylene respo	1.39e+01
c	42	20	15.6	5746 24	T09225 Partial sequence of v	1.39e+01
c	43	20	15.6	133894 17	T13635 ACNPV genomic DNA clo	1.39e+01
c	44	20	15.6	580073 27	T58640 Mycoplasma genitalium	1.39e+01

ALIGNMENTS

RESULT 1
 ID T20320 standard; cDNA to mRNA: 128 BP.

AC T20320;
 AT 18-JUL-1996 (first entry)
 DE Human gene signature HUMGS01470.

KM Gene signature; messenger RNA: mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 OS Homo sapiens.
 KW cell typing; abnormal cell function; ss.

PN W09514772-A1.
 PD 01-JUN-1995.

PF 11-NOV-1994; J01916.
 PR 12-NOV-1995; JP-355504.

PA (MATSU) MATSUBARA K.
 PI (OKUBO) OKUBO K.

PI Matsubara K, Okubo K;
 DR WPI: 95-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 for diagnosis of abnormal cell function, by preparing cDNA that
 reflects relative abundance of corresp. mRNA in specific human
 tissues

PS Claim 1: Page 612; 2245pp; Japanese.
 CC A single-stranded DNA (or its complementary strand or the corresp.
 double-stranded DNA) which comprises one of the 7837 "GS" sequences
 given in T19001-T26837 and which is able to hybridize to part of
 human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 sequences were obtained from 3'-directed cDNA libraries prepared
 from various human tissues; synthesis of cDNA was initiated from the
 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 untranslated sequence is unique to a particular mRNA species, almost
 all the 3'-oriented cDNAs hybridize with specific mRNAs. Each library
 is constructed so as to reflect accurately the relative abundance of
 different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 determined (esp. using primers and probes derived from the GS
 sequences) as a means of diagnosing abnormal cell function or for
 recognizing different cell types.

```

SO Sequence      128 BP;          41 A;         22 C;         18 G;        47 T;

Query Match
Best Local Similarity 100.0%; Score 128; DB 19; Length 128;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dd   1 gatagaattttaaacgagaatttgcacaaaggcatttttgtctacccttla 60
      |||||||
Oy   1 GATAGAAATTTTAAACGACAATTTGTCAAAAGGCATTTTTTGTCACCCCTTTA 60

Db   61 cacttttcagattccaaagtgtctcatctcaacctttaaagaataaatatcttc 120
      |||||||
Oy   61 CACTTTTCAGATTCTCAAAGTGTCTCATCTCAACTTTTAAAGAATAAAGATATCTTC 120

Db   121 tggcgcaa 128
      |||||||
Oy   121 TGGCGCAA 128

RESULT 2
ID O51746 standard; cDNA; 91 BP.
AC O51746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PR 24-MAY-1993; 108325.
PF 26-MAY-1992; US-889651.
PA (BECT ) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PT Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (O51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also O51735-45 and O51747-59.
CC Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;
SQ

Query Match              27.3%; Score 35; DB 9; Length 91;
Best Local Similarity 0.0%; Pred. NO. 2.38e-06;
Matches 0; Conservative 41; Mismatches 6; Indels 0; Gaps 0;

Db   13 vhsyyvvhvshsvhvhhvvhvsvvvvvhvvhvvhvhyhyys 59
      :::::
Cp   63 GGTAAAGGGGTAGACAAAAGAACCTTTTGACAAAATTCTGC 17

RESULT 3
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc-feature 19..69
FT FT //tag= a
FT primer_bind //function=multiple cloning site
FT FT 187..204
FN EP-285123-A. /tag= b
FN PD 05-MAY-1988.
PR 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
LEhtovaara P, Knowles J, Koivuola A, Bamford J, Reinikainen T;
```

```

PT      Introducing random point mutations into nucleic acids -  

PR      WPI; 88-279927/40.  

PS      By prepn of single stranded template, annealing a primer, elongation,  

PT      misincorporation, completion of molecules and screening.  

PS      Disclosure; P; English.  

CC      Random point mutations were introduced into the alpha fragment of  

CC      E.coli beta-galactosidase. The wild type sequence was obtained as a  

CC      single stranded template and an oligonucleotide was hybridised to  

CC      it to generate a popn of DNA molecules which terminate at all  

CC      possible nucleotide positions within a specified region. The  

CC      variable 3' ends generated in this way are used as primers for  

CC      reverse transcriptase. Nucleotides are misincorporated by the  

CC      transcribtase and the molecules are completed to forms that can be  

CC      amplified and then expressed in a suitable host-vector system.  

CC      The sequence covers all 176 given base substitutions, most of which  

CC      occurred singularly in any diff mutant.  

CC      See also P80575.  

SQ      Sequence   204 BP;    21 A;   47 C;   17 G;   11 T;   108 Others;  

Query Match          27.3%; Score 35; DB 1; Length 204;  

Best Local Similarity 5.0%; Pred. No. 2,38e-06;  

Matches           5; Conservative       56; Mismatches 40; Indels     0; Gaps     0;  

Db        85 gymtcthyrmbvnyrdynrsdaawcyocryskyacgnachdhvybbbvya 144  

Qy        17 GCAGAAATTTCTCAAAAAGGGCATATTTTGTGTCACCCCTTTCACACTTTCAGATITC 76  

         ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  

Db        145 vhhnnncccbhvhcgvhbhnbmhmwayvrhdarrddvh 185  

Qy        77 AAAGTGCTCATCTCACTCACTTTTAAGAATAAAGATATCT 117  

         ::||::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  

RESULT      4  

ID      Q51746 standard; cDNA; 91 BP.  

AC      Q51746;  

DT      31-MAY-1994 (first entry)  

DE      Oligonucleotide probe MK14-A  

KW      Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;  

KM      ss.  

OS      Synthetic.  

PN      EP-571911-A.  

PD      01-DEC-1993.  

PF      24-MAY-1993; 108325.  

PR      26-MAY-1992; US-889651.  

PA      (BECT ) BECTION DICKINSON CO.  

PI      Shank DD, Spears PA;  

DR      WPI; 93-378844/48.  

PT      New oligo:nucleotide probes specific for Mycobacteria - used for  

PT      detection and amplification of Mycobacteria nuclei acid in  

PT      samples  

PS      Claim 3; Page 14; 23pp; English.  

CC      Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14  

CC      (Q51733). It hybridized to all spp. of mycobacteria tested, but  

CC      cross reacted to a few non-mycobacterial spp. The probe may  

CC      be useful as an initial screen for mycobacterial infection.  

CC      See also Q51735-45 and Q51747-59.  

SQ      Sequence   91 BP;      5 A;   17 C;   15 G;    4 T;  

Query Match          22.7%; Score 29; DB 9; Length 91;  

Best Local Similarity 4.1%; Pred. No. 1,61e-03;  

Matches           2; Conservative       37; Mismatches 10; Indels     0; Gaps     0;  

Db        14 hasyvvhvshhsvhvvhvhsvvvvhvvhvvhvhyvsvct 62  

Qy        73 TCACAAGTGCTCATCTCACTTTTAAGAATAAAGATATCTGCT 121  

         :::::||||::::|:::|:::|:::|:::|:::|:::|:::|:::|:::  

RESULT      5  

ID      N81164 standard; DNA; 204 BP.  

AC      N81164;  

DT      08-NOV-1990 (first entry)  

DE      Base substituted E.coli beta-galactosidase alpha-fragment.  

KW      E.coli beta galactosidase alpha-fragment; base substitutions; ss.
```


FH	Escherichia coli.
FT	Key Location/Qualifiers
FT	misc_feature 19..69
FT	/tag= a
FT	/function-multiple cloning site
FT	primer_bind 187..204
FT	/tag= b
PB	EP-285123-A.
PD	05-MAY-1988.
PF	30-MAR-1988; 105163.
PR	03-APR-1987; US-034819.
PA	(SUSO) SUOMEN SOKERI OY.
PI	Lehtovaara P, Knowles J, Koivuola A, Bamford J, Reinikainen T;
PT	WPI: 88-279927/40.
PS	Introducing random point mutations into nucleic acids - by prepn of single stranded template, annealing a primer, elongation, PT minincorporation, completion of molecules and screening.
CC	Disclosure: P: English.
CC	Random point mutations were introduced into the alpha fragment of E.coli beta-galactosidase. The wild type sequence was obtained as a CC single stranded template and an oligonucleotide was hybridised to it to generate a popn of DNA molecules which terminate at all CC possible nucleotide positions within a specified region. The CC variable 3' ends generated in this way are used as primers for reverse transcriptase. Nucleotides are misincorporated by the CC transcriptase and the molecules are completed to forms that can be amplified and then expressed in a suitable host-vector system.
CC	The sequence covers all 176 diff base substitutions, most of which occurred singularly in any given mutant.
CC	See also P80575.
SQ	Sequence 204 BP: 21 A: 47 C: 17 G: 11 T: 108 Others;
Query Match	20.3%; Score 26; DB 1; Length 204;
Best Local Similarity	6.5%; Pred. No. 3.70e-02;
Matches	6; Conservative 49; Mismatches 37; Indels 1; Gaps 1;
Dc	99 dnvyrdyrdsaaawcyccyrrsvkyddccymachdhdyvbybvnyvhnhnncccbbn 138 ::: : : : : : : : : : : : : : : : : : :
Cp	125 GCCCGAGCAGATATTCTTATTCCTTTAAAGGTGAG-ATGAGACACTTGAGNATCTGA 67
Db	159 hvchnvbnhnrmwayvrhdarddvhhvcgbc 191 ::: : : : : : : : : : : : : : : : : : :
Cp	66 AAAGTGTAAGAAGGGGTAGCAAAAAAATGACC 34
RESULT	6
ID	O70473 standard; DNA; 114 BP.
AC	O70473:
DT	10-APR-1995 (first entry)
KM	Generic DNA sequence to generate a random TSAR peptide library.
KW	TSAR; totally synthetic affinity reagent; sythetic; binding domain;
KW	effector domain; concatenated heterofunctional protein; linker;
OS	direct; rapid; detection; screening; treatment; generic; ss.
SY	Synthetic.
FT	Key Location/Qualifiers
FT	misc_feature 55..60
FT	/tag= a
FT	/note= "encoded by Z"
PN	MO9418318-A.
PD	18-AUG-1984.
PE	01-FEB-1984; U00977.
PR	01-FEB-1993; US-013416.
PR	30-DEC-1993; US-176500.
PR	31-JAN-1994; US-189331.
PA	(UYNC-) UNIV NORTH CAROLINA.
PI	Fowlkes DM, Kay BK,
PT	WPI: 94-279739/34.
DR	P-PSTD: R58393.
PT	Identifying proteins or peptide(s) which bind a ligand - by PT screening a recombinant vector library expressing fusion proteins
PT	comprising a binding domain and an effector domain
DS	Disclosure: Page 36, 25pp: English.
CC	O70473 is a generic DNA sequence used to generate random TSAR (Totally

[illegible]

CP 60 TAA 58

RESULT 11
ID 027945 standard: DNA: 2190 BP.
AC 027945:
DE 28-JAN-1993 (first entry)
DT Sequence of CryIIIC(b) gene.
KW Insecticidal crystal protein: coleoptera; Colorado potato beetle;
KM Diabrotica: 70KD insecticidal protein: CryIII gene; ss.
OS Bacillus thuringiensis strain EG5144.
FH Key
FT rbs
FT 132...137
FT cds
FT 144...2102
FT /*tag= a
FT /*tag= b

WO9213954-A.
PD 20-AUG-1992.
PE 03-JAN-1992: U000040.
PR 31-JAN-1991: US-649562.
PR 23-DEC-1991: US-813592.
PA (ECOG-) ECGEN INC.
PI Donovan WP, Rupar MJ, Slaney AC;
DR WPI: 92-300040/36.
DR P-PSDB: R26574.
PT Bacillus thuringiensis contg. CryIIIC (b) gene in plasmid - used
PT to make insecticide against coleopteran insects
PS Claim 1: Fig 1: 91pp: English.
CC A comparison of the nucleotide base sequence of the CryIIIC(b) gene
CC coding region with the corresp. coding region of the prior art
CC CryIIIA gene indicates significant differences between the two
CC genes. The CryIIIC(b) gene is only 76% homologous (positionally
CC identical) with the CryIIIA gene. The sequence of the CryIIIC(b)
CC gene coding region with the corresp. coding region of the CryIIIB
CC gene obtained from recently discovered B.t. strain EG2838 (NRRL
CC Accession No. B-18603) indicates that the CryIIIC(b) gene is 96%
CC homologous (positionally identical) with the CryIIIB gene.
SO Sequence 2190 BP: 786 A; 329 C; 411 G; 664 T;

Query Match 18.0%; Score 23; DB 4; Length 2190;
Best Local Similarity 68.3%; Pred. No. 7.67e-01;
Matches 43; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db 1267 gcaagatatttggtgtacgaagttgatttgcataatgatgtcctcaaaaatg 1326
CP 120 GCAAGATATTCTTTATTCTTTAAAGTTGAGATGACACTTGTGAAATCGAANAAGTG 61

Db 1327 aaa 1329
CP 60 TAA 58

RESULT 12
ID 005679 standard: DNA: 2425 BP.
AC 005679:
DE 03-JAN-1991 (first entry)
DT btrPGS1208 gene.
KM Toxin: crystal; insecticide; ss.
OS Bacillus thuringiensis strain PGS1208.
FH Key
FT cds
FT 342...2297
FT /*tag= a
FT /*tag= b
FT /*product=74 KD BtrPGS1208 protoxin
FT 513...2294
FT /*tag= b
FT /*product=67 KD BtrPGS1208 toxin

EP-382990-A.
PD 22-AUG-1990.
PR 15-FEB-1989: 400428.
PR 15-FEB-1989: EP-400428.
PA (PLAN-) PLANT GENETIC SYST.
PI Peferoen M, Lambert B, Joos H;
WPI: 90-255362/34.

DR P-PSDB: R06460.
PT New Bacillus thuringiensis strains - producing toxin active
PT against Coleoptera
PS Claim 3: Fig 1: 30pp: English.
CC The BtrPGS1208 strain was isolated from grain dust and was deposited
CC at the DSM (19/1/1989) under Acc. No. 5131. The protoxin crystals
CC produced on expression of the gene can be digested with trypsin to
CC produce the mature protein. Plants contg. the gene are resistant
CC to Coleoptera pests while the crystals are useful as insecticides.
CC The gene can also be truncated and used to produce chimaeric genes
CC together with the btrPGS1245 gene (Q05680).
SO Sequence 2425 BP: 872 A; 366 C; 429 G; 758 T;

Query Match 18.0%; Score 23; DB 1; Length 2425;
Best Local Similarity 68.3%; Pred. No. 7.67e-01;
Matches 43; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db 1585 gcaagatatttggtgtacgaagttgatttgcataatgatgtcctcaaaaatg 1644
CP 120 GCAAGATATTCTTTATTCTTTAAAGTTGAGATGACACTTGTGAAATCGAANAAGTG 61

Db 1645 aaa 1647
CP 60 TAA 58

RESULT 13
ID T13613 standard: DNA: 74 BP.
AC T13613:
DE 15-OCT-1996 (first entry)
DT DC43 TSAR library generating oligonucleotide #4.
KW Cis-element; ligand; highly specific DNA binding domain; HSDB; human;
KW NF-kappa-B; regulatory domain; murine; homologue; IL-6; interleukin;
KW HLA class I gene; HLA-B*7; HLA-J; HLA-A*2; acute phase protein response;
KW syngene; immune response; IL-8; IL-16; diagnosis; gene therapy;
KW transcription; ds.
OS Synthetic.
FH Key
FT misc_feature
FT 6..65
FT /*tag= a
FT /*note= "Random peptide generating fragment"

WO9606188-A1.
PD 29-FEB-1996.
PR 17-AUG-1995: U10523.
PR 18-AUG-1994: US-292902.
PR 15-AUG-1995: US-515190.
PA (CYTO-) CYTOGEN CORP.
PI Fowlkes DM, Rodwell JD;
WPI: 96-151391/15.
PT Synthetic random nucleotide sequences encoding ligand binding
PT domains - identified by screening library of vectors or peptide(s)
PT and useful for gene therapy and diagnosis
PS Disclosure: Page 15: 224pp: English.
CC The sequences given in T13610-13 represent oligonucleotides which
CC were used in the generation of the DC43 TSAR random peptide library.
CC The sequence of amino acids encoded by these sequences are covered by
CC the generic sequence given in W00125. They are unpredictable and
CC substantially random in sequence. All 20 naturally occurring amino acids
CC are encoded by the variable coding region. Invariant nucleotides are
CC positioned at particular sites in these nucleotide sequences to aid the
CC assembly and cloning of the synthesised oligonucleotides. At the 5'
CC termini of each set of the oligonucleotides, the invariant nucleotides
CC encode for efficient restriction endonuclease cleavage sites. The
CC invariant nucleotides at the 5' termini are chosen to encode pairs of
CC sites for by restriction enzymes which can function in the same buffer
CC conditions, are commercially available at high specific activity, are
CC not complementary to each other to prevent self-ligation of the
CC synthesised double stranded oligonucleotides and which require 6 or 8
CC nucleotides for a cleavage recognition site in order to lower the
CC frequency of cleaving within the inserted double stranded synthesised
CC oligonucleotide sequences. The 3' termini invariant nucleotide
CC positions are complementary pairs of 6, 9 or 12 nucleotides to aid in
CC annealing two synthesised single stranded sets of nucleotides together

[illegible]

```

Db      72 nacc 75
      11
Cp      38 GACC 35

RESULT 15
ID      070468 standard; DNA; 114 BP.
AC      070468;
DT      05-APR-1995 (first entry)
DE      Generic DNA sequence to generate a random TSAR peptide library.
KW      TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW      effector domain; concatenated heterofunctional protein; linker;
KW      direct; rapid; detection; screening; treatment; generic; ss.
OS      Synthetic.
FH      Key
FT      misc.feature
FT      Location/Qualifiers
FT      55..60
FT      /tag= a
FT      /note= "this sequence represents 'Z'; Z can be a
FT      sequence of 6, 9 or 12 nucleotides (see
FT      comments)"
FN      W09418318-A.
FN      18-AUG-1994.
FN      01-FEB-1994; U00977.
FN      01-FEB-1993; US-013416.
FN      30-DEC-1993; US-176500.
FN      31-JAN-1994; US-189331.
FN      (UYNC) UNIV NORTH CAROLINA.
FN      Fowlkes DM, Kay BK;
FN      WPI; 94-279739/34.
FN      P-PSDB; R65154.
FN      Identifying proteins or peptide(s) which bind a ligand - by
FN      screening a recombinant vector library expressing fusion proteins
FN      comprising a binding domain and an effector domain
PS      Disclosure; page 35, 255pp; English.
CC      Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC      Synthetic Affinity Reagents) peptides. This generic formula can also be
CC      represented as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC      and Y are flanking restriction sites (X is not the same as Y) that are
CC      not specified further. Other generic sequences are shown in Q70466-68.
CC      Other specific peptides generated by these generic sequences are shown in
CC      R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC      comprising at least two functional regions - a binding domain with
CC      affinity for a ligand and a second effector peptide portion that is
CC      chemically or biologically active. They may further comprise a linker
CC      peptide between the 2 domains. The oligonucleotides are also designed so
CC      that the expressed peptide contains 2 or 4 cysteine residues positioned
CC      in, or flanking, the unpredicted or variant residues. These residues
CC      confer some degree of conformational rigidity to the peptides. The TSARs
CC      or comps. comprising a TSAR binding domain can be used in vivo to
CC      deliver a chemically or biologically active moiety, eg. metal ion,
CC      radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC      cell. They can also replace the function of macromolecules, eg.
CC      monoclonal or polyclonal antibodies and therefore circumvent the need
CC      for complex methods of hybridoma formation or in vivo antibody
CC      production. The TSARs are easily characterised and have designed activity
CC      allowing direct and rapid detection in a screening process.
SQ      Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 17.2%; Score 22; DB 12; Length 114;
Best Local Similarity 3.08; Pred. NO. 2.05e+00;
Matches 3; Conservative 26; Mismatches 71; Indels 0; Gaps 0;

Db      9 bmbnbmbnbmbnbmbnbmbnbtgcmbnbmbnbmbnbmbnbmbnnnnbmbnbmbn 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      25 TTTCTCAAAAGGCGCATTTTGTGTCTACCCCTTTACACTTTTCAGATCTCTCAAGGTC 84
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      69 bmbnbmbnbmbnbtgcmbnbmbnbmbnbmbnbmbnbmbnmb 108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      85 TCATCTCAACTTTTAAAGAAATAAAGATATATCTTGCTGGG 124

```

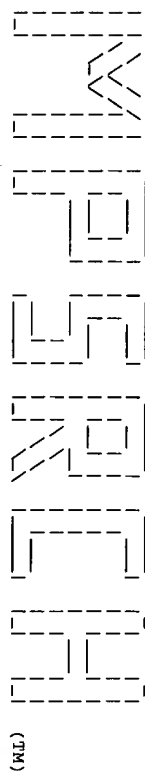
Search completed: Tue Dec 22 04:58:00 1998

Tue Dec 22 09:17:10 1998

US-08-530-112A-1320.rng

Page 7

Job time : 41 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 04:53:05 1998; MasPar time 228.43 Seconds

Tabular output not generated. 1003.437 Million cell updates/sec

Title: >US-08-530-112A-1320

Description: (1-128) from US08530112A.seq

Perfect Score: 128

N.A. Sequence: 1 GATCGAATTTTAAAGCAG.....ACATATCTTGCTGGGCAAA 128

Comp: CTAGCTTAAATAATTCGTC.....TCCTATAGAACGACCCGTTT

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 segs, 895388244 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

embl-est15

Database:

genbank-est107

5:gb-est1 6:gb-est10 7:gb-est11 8:gb-est12 9:gb-est13

10:gb-est14 11:gb-est15 12:gb-est16 13:gb-est17

14:gb-est18 15:gb-est19 16:gb-est20 17:gb-est21

18:gb-est22 19:gb-est23 20:gb-est24 21:gb-est25 22:gb-est26

23:gb-est27 24:gb-est28 25:gb-est29 26:gb-est30 27:gb-est31

28:gb-est32 29:gb-est33 30:gb-est34

Statistics: Mean 9.501; Variance 2.560; scale 3.711

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	128	100.0	128	7	HUMGS01470	Human HL60 3'directed	1.62e-147
2	93	72.7	401	19	H78523	YU13C02.r1 Homo sapien	7.68e-96
3	79	61.7	380	21	AA053439	2170B05.r1 Striatogene	1.08e-75
4	33	25.8	252	12	AA754459	97SN1787 Rice Immature	1.89e-14
5	32	25.0	252	12	AA754459	97SN1787 Rice Immature	2.72e-13
6	28	21.9	2275	11	AF034173	Homo sapiens rnc2 co	8.32e-09
7	26	20.3	428	13	AA867655	vx12e03.r1 Soares 2NDM	1.14e-05
8	25	19.5	247	12	AA754458	97SN1784 Rice Immature	1.24e-05
9	25	19.5	446	10	AA394768	26565 Lambda-PRU2 Arab	1.24e-05
10	24	18.8	300	9	C57753	C.elegans CDNA clone y	1.29e-04
11	24	18.8	300	9	C55145	C.elegans CDNA clone y	1.29e-04
12	24	18.8	300	9	C53653	C.elegans CDNA clone y	1.29e-04
13	24	18.8	300	9	C55340	C.elegans CDNA clone y	1.29e-04

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
14	24	18.8	349	6	CELK034GYR	C.elegans CDNA clone y	1.29e-04
15	24	18.8	368	6	CELK032F4R	C.elegans CDNA clone y	1.29e-04
16	24	18.8	379	6	CELK032B1R	C.elegans CDNA clone y	1.29e-04
17	24	18.8	390	6	CELK024HXR	C.elegans CDNA clone y	1.29e-04
18	24	18.8	399	6	CELK043G9R	C.elegans CDNA clone y	1.29e-04
19	24	18.8	430	7	AA607659	zx60h04.r1 Soares test	1.29e-04
20	23	18.0	240	9	C31634	C.elegans CDNA clone y	1.26e-03
21	23	18.0	300	9	C58196	C.elegans CDNA clone y	1.26e-03
22	23	18.0	300	9	C30068	C.elegans CDNA clone y	1.26e-03
23	23	18.0	339	28	AQ097886	HS_3036.B2.A04_MF CIT	1.26e-03
24	23	18.0	407	26	AQ010078	HS_2263.B2.MR.C05 CIT	1.26e-03
25	23	18.0	446	5	R26820	Y44h12.s1 Homo sapien	1.26e-03
26	23	18.0	2275	11	AF034173	Homo sapiens rnc2 co	1.26e-03
27	22	17.2	247	12	AA754458	97SN1784 Rice Immature	1.15e-02
28	22	17.2	276	20	W96523	ze37g01.r1 Soares ret1	1.15e-02
29	22	17.2	311	13	AA644269	at63d05.s1 Soares Nhm	1.15e-02
30	22	17.2	315	13	AA629430	ae53h05.s1 Striatogene	1.15e-02
31	22	17.2	394	8	AA257038	zr84h10.s1 Soares Nhm	1.15e-02
32	22	17.2	396	13	AA812038	yb67f06.s1 NCI_CGAP_GC	1.15e-02
33	22	17.2	427	19	N68202	Y737e12.s1 Homo sapien	1.15e-02
34	22	17.2	434	19	N52622	OS23d04.s1 NCI_CGAP_K1	1.15e-02
35	22	17.2	439	15	AA989415	CIT-HSP-2028N9.TF CIT	1.15e-02
36	22	17.2	464	27	B74176	aa55c12.r1 NCI_CGAP_GC	1.15e-02
37	22	17.2	508	8	AA488937	z495c08.r1 Soares tota	1.15e-02
38	22	17.2	518	8	AA476805	ze37g12.r1 Soares ret1	1.15e-02
39	22	17.2	519	22	AA017088	n203h05.s1 NCI_CGAP_AA	1.15e-02
40	22	17.2	543	12	AA470880	n741d07.s1 NCI_CGAP_GC	1.15e-02
41	22	17.2	548	9	AA579862	z68h02.s1 Striatogene	1.15e-02
42	22	17.2	560	22	AA205666	ok65h05.s1 NCI_CGAP_GC	1.15e-02
43	22	17.2	634	14	AA910648	RC111-53L13.TJ RPl111	1.15e-02
44	22	17.2	636	27	AQ051866	me74g10.r1 Soares mous	1.15e-02
45	22	17.2	945	10	W12977		1.15e-02

ALIGNMENTS

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	128	100.0	128	7	HUMGS01470	Human HL60 3'directed MboI CDNA, HUMGS01470, clone pm1222, mRNA sequence.	17-JUN-1996
2	93	72.7	401	19	H78523	YU13C02.r1 Homo sapien	7.68e-96
3	79	61.7	380	21	AA053439	2170B05.r1 Striatogene	1.08e-75
4	33	25.8	252	12	AA754459	97SN1787 Rice Immature	1.89e-14
5	32	25.0	252	12	AA754459	97SN1787 Rice Immature	2.72e-13
6	28	21.9	2275	11	AF034173	Homo sapiens rnc2 co	8.32e-09
7	26	20.3	428	13	AA867655	vx12e03.r1 Soares 2NDM	1.14e-05
8	25	19.5	247	12	AA754458	97SN1784 Rice Immature	1.24e-05
9	25	19.5	446	10	AA394768	26565 Lambda-PRU2 Arab	1.24e-05
10	24	18.8	300	9	C57753	C.elegans CDNA clone y	1.29e-04
11	24	18.8	300	9	C55145	C.elegans CDNA clone y	1.29e-04
12	24	18.8	300	9	C53653	C.elegans CDNA clone y	1.29e-04
13	24	18.8	300	9	C55340	C.elegans CDNA clone y	1.29e-04

REFERENCE 1 (sites)
Okubo, K., Itoh, K., Fukushima, A., Yoshii, J. and Matsubara, K.
Monitoring cell physiology by expression profiles and discovering
cell type-specific genes by compiled expression profiles
Genomics 30 (2), 178-186 (1995)

JOURNAL
TITLE
AUTHORS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Submitted (10-Sep-1993) to the DDBJ/EMBL/GenBank databases. Kousaku
Okubo, Osaka University, Institute for Molecular and Cellular Bio:
1-3, Yamada-oka, Suita, Osaka Pref. 565, Japan
(E-mail:kousaku@imcb.osaka-u.ac.jp, Tel:06-877-5111(ex.3315),
Fax:06-877-1922)
Submitted (10-Sep-1993) to DDBJ by:
Kousaku Okubo
Institute for Molecular and Cellular Biology

Osaka University
3-1, Yamadaoka
Suita, Osaka, 565

Japan
Phone: 06-877-5111
Fax : 06-875-1922.

FEATURES
source
Location/Qualifiers
1. .128
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="H160"
/cell_type="promyelocyte"
/dev_stage="adult"
/sex="female"

BASE COUNT
41 a 22 c 18 g 47 t

ORIGIN

Query Match 100.0%; Score 128; DB 7; Length 128;
Best Local Similarity 100.0%; Pred. No. 1,62e-147;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GATCAGATTTTAAAGCAGATTTTGTCAAAGGTCATTTTGTCTACCCCTTTA 60
|||||
Qy 1 GATCAGATTTTAAAGCAGATTTTGTCAAAGGTCATTTTGTCTACCCCTTTA 60
|||||
Db 61 CACTTTTCAGATTCACAAGTGTCTCATCTCAACTTTTAAAGATTAAGATATCTTGC 120
|||||
Qy 61 CACTTTTCAGATTCACAAGTGTCTCATCTCAACTTTTAAAGATTAAGATATCTTGC 120
|||||
Db 121 TGGGCAAA 128
|||||
Qy 121 TGGGCAAA 128

RESULT 2
LOCUS H78523 401 bp mRNA EST 09-NOV-1995
DEFINITION YU13C02.r1 Homo sapiens cDNA clone 233666 5'
ACCESSION H78523
NID 91056612
KEYWORDS EST.
SOURCE human clone-233666 (primer-M13R1 library-Soares fetal liver spleen
INPIS vector-PT73D (Pharmacia) with a modified polylinker
host-DH10B (ampicillin resistant) Rstrel-Pac I Rstrel2-Eco RI liver
and spleen from a 20 week post conception male fetus. 1st strand
cDNA was primed with a Pac I - Oligo(dT) primer [5'
AAGTGAAGATTAATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
I and cloned into the Pac I and Eco RI sites of the modified pT73
vector. Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 401)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

TITLE The Washu-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT

Contact: Wilson RK
Washu-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 340
Source: IMAGE Consortium, LNL.

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.

FEATURES
source
Location/Qualifiers
1. .401
/organism="Homo sapiens"
/clone="233666"

BASE COUNT
104 a 68 c 60 g 167 t 2 others

Query Match 72.7%; Score 93; DB 19; Length 401;
Best Local Similarity 98.1%; Pred. No. 7.68e-96;
Matches 105; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Db 283 GATCAGATTTTAAAGCAGATTTTGTCAAAGGTCATTTTGTCTACCCCTTTT 341
|||||
Qy 1 GATCAGATTTTAAAGCAGATTTTGTCAAAGGTCATTTTGTCTACCCCTTTT 59
|||||
Db 342 ACACCTTTTCAGATTCACAAGTGTCTCATCTCAACTTTTAAAGAT 388
|||||
Qy 60 ACACCTTTTCAGATTCACAAGTGTCTCATCTCAACTTTTAAAGAT 106
|||||

RESULT 3
LOCUS AA053439 380 bp mRNA EST 13-SEP-1996
DEFINITION Z170B05.r1 Stratagene colon (#937204) Homo sapiens cDNA clone
509937 5' similar to gb:X02492 INTERFERON-INDUCED PROTEIN 6-16
PRECURSOR (HUMAN); mRNA sequence.
ACCESSION AA053439
NID 91544094
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; mitochondria eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 380)

REFERENCE
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,
Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE Washu-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT

Contact: Wilson RK
Washu-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 342.
Location/Qualifiers
1. .380
/organism="Homo sapiens"
/note="Organ: colon; Vector: pBluescript SK-; Site:1;
EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. T-84 colonic epithelial cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3'."
/db_xref="taxon:9606"
/clone="509937"
/clone_id="Stratagene colon (#937204)"
/lab_host="SOLR cells (kanamycin resistant)"
1. .380

FEATURES
source
Location/Qualifiers
1. .380

BASE COUNT
114 a 79 c 80 g 104 t 3 others

Query Match 61.7%; Score 79; DB 21; Length 380;
Best Local Similarity 100.0%; Pred. No. 1.08e-75;

Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 TCTACCCCTTACCTTTCAGATTCATCAAGTCTCATCTCACTTTAAAGAATA 60
|||||
Oy 48 TCTACCCCTTACCTTTCAGATTCATCAAGTCTCATCTCACTTTAAAGAATA 107
|||||

Db 61 AAGAATATCTTGTGGCA 79
|||||

Oy 108 AAGAATATCTTGTGGCA 126

RESULT 4
LOCUS AA754459 252 bp mRNA EST 20-JAN-1998
DEFINITION 97SN187 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
ACCESSION AA754459
NID 92801165
KEYWORDS EST.
SOURCE rice.
ORGANISM Oryza sativa
Eukaryote; Viridiplantae; Charophyta/Embryophyta group;
Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;
Poales; Poaceae; Oryza.
1 (bases 1 to 252)
Nahm,B.H., Kim,W.Y., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
Lee,M.C. and Eun,M.Y.
Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
Unpublished (1998)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CONTACT: Eun M.Y.
Department of Cyto Genetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeungsun20.astl.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea, 449-728 bhnahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1..252
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/clone="97SN187"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT 5 a 21 c 12 g 35 t 179 others
ORIGIN

Query Match 25.8%; Score 33; DB 12; Length 252;
Best Local Similarity 7.1%; Pred. No. 1,89e-14;
Matches 8; Conservative 61; Mismatches 43; Indels 1; Gaps 1;

Db 20 TTSYBCHGNBWWCVASHGNVMSVHNCIBRGHCCKKNVMTKMGTYNMENVSQDMHY 79
|||||

Cp 127 TTGCCACCAAGATATCTTTATTTTAAAGTTGAGATG-AGACACTTTGAGAAATCT 69
|||||

Db 80 WBNVTKDVGNHTRCSRWBRTRAHYHDYNCBYYNNNDYHMMHBBMYBBT 132
|||||

Cp 68 GAAAGCTTAAGAGCGGTAGACAAAATAATGACCTTTGACAAATCTGCT 16
|||||

RESULT 5
LOCUS AA754459 252 bp mRNA EST 20-JAN-1998
DEFINITION 97SN187 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
ACCESSION AA754459
NID 92801165
KEYWORDS EST.
SOURCE rice.
ORGANISM Oryza sativa
Eukaryote; Viridiplantae; Charophyta/Embryophyta group;
Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;
Poales; Poaceae; Oryza.
1 (bases 1 to 252)
Nahm,B.H., Kim,W.Y., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
Lee,M.C. and Eun,M.Y.
Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
Unpublished (1998)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CONTACT: Eun M.Y.
Department of Cyto Genetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeungsun20.astl.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea, 449-728 bhnahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1..252
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/clone="97SN187"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT 5 a 21 c 12 g 35 t 179 others
ORIGIN

Query Match 25.0%; Score 32; DB 12; Length 252;
Best Local Similarity 14.6%; Pred. No. 2.72e-13;
Matches 12; Conservative 41; Mismatches 29; Indels 0; Gaps 0;

Db 83 BNTKDVGNHTRCSRWBRTRAHYHDYNCBYYNNNDYHMMHBBMYBBTGMCMC 142
|||||

Oy 10 TTTTAAGCAGATTTTGTGCAAAAGGTCATTTTGTGTACCTTTTACACTTTCA 69
|||||

Db 143 BHNTRCTASGWHSTNYDKS 164
|||||

Oy 70 GATTCCAAGTGTCTCATCTC 91
|||||

RESULT 6
LOCUS AF034173 2275 bp mRNA EST 22-DEC-1997
DEFINITION Homo sapiens ntcon2 contig mRNA, partial sequence, mRNA sequence.
ACCESSION AF034173
NID 92707735
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 2275)
Tripodis,N. and Ragousstis,J.
Generation of a transcription map in the region immediately
centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
boundary
Unpublished
2 (bases 1 to 2275)
Tripodis,N. and Ragousstis,J.
Direct Submission

JOURNAL Submitted (13-NOV-1997) Division of Medical and Molecular Genetics,
Guy's Hospital, 7th floor, Guy's Tower, London SE1 9RT, UK

FEATURES
Source
1. 2275
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
/clone="ntcon2 contig"
/tissue_type="liver; brain"
/dev_stage="fetus"
/note="Similar to Br140"
BASE COUNT 438 a 619 c 470 g 599 t 149 others
ORIGIN

Query Match 21.9%; Score 28; DB 11; Length 2275;
Best Local Similarity 12.7%; Pred. No. 8,32e-09;
Matches 14; Conservative 60; Mismatches 34; Indels 2; Gaps 2;

Db 1506 MNAMCAMAACWYWMKMKCKYRKYKYYSTYKSNRMYWT-YYWYCWCTTS 1564

QY 14 AAGAGAGATTTTGCAAAAGGTCATTTTGTCTACCCCTTTACACTTTTCAGATT 73

Db 1565 MNSACAMRNMGYMSRSRSYWGWSGCGMTKRYRYSMTGWTK 1614

QY 74 CTCGA-AGTCTCATCTCACTTTTAAAGATTAAGATATCTTGCTG 122

RESULT 7 AA867655 428 bp mRNA EST 16-MAR-1998

LOCUS vx12e03.r1 Soares 2NDMT Mus musculus cDNA clone 1264252 5', mRNA
DEFINITION sequence.

ACCESSION AA867655

NID 92963100

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

AUTHORS Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 428)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Schlesinger,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Waterson,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterson,R.

The WashU-HM Mouse EST Project

Unpublished (1996)

TITLE

JOURNAL

COMMENT

Contact: Marra M/Mouse EST Project

WashU-HM Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:666804

Seg primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 427.

Location/Qualifiers

1. 428

/organism="Mus musculus"

/strain="C57BL/6J"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5',

TGTACCAATCTGAACTGGAGCGCCGCTTTTCTTTTCTTTTCTTTTCTTTT

3'] ; double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT7T3 vector. RNA

provided by Dr. Bertrand Jordan. Library went through two

rounds of normalization, and was constructed by Bento

Soares and M.Fatima Bonaldo."

/db_xref="taxon:10090"

/clone="1264252"

/clone_1lb="Soares 2NDMT"

/sex="male"

/tissue_type="thymus"

/dev_stage="4 weeks"

/lab_host="DH10B"

/lab_host="DH10B"

BASE COUNT 134 a 71 c 97 g 124 t 2 others

ORIGIN

Query Match 20.3%; Score 26; DB 13; Length 428;

Best Local Similarity 75.0%; Pred. No. 1.14e-06;

Matches 39; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Db 170 TTTAGAGACTCCCTAAGGGCTCATCTTAACCTTTAAAGAAAGAAAT 221

QY 64 TTTTCAGATTCGAAAGTGTCTCATCTCAACTTTTAAAGATTAAGATAT 115

RESULT 8 AA754458 247 bp mRNA EST 20-JAN-1998

LOCUS 97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa

DEFINITION cDNA clone 97SN1784, mRNA sequence.

ACCESSION AA754458

NID 92801164

KEYWORDS EST.

SOURCE rice.

ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Charophyta/Embryophyta group;

Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;

Poales; Poaceae; Oryza.

1 (bases 1 to 247)

Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,

Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,

Lee,M.C. and Eun,M.Y.

large-scale sequencing analysis of ESTs from Rice Immature Seed

Unpublished (1998)

TITLE

JOURNAL

COMMENT

Contact: Eun M.Y.

Department of CytoGenetics

National Inst. of Agric. Sci. and Tech, RDA

Tel: 82 331 290 0301

Fax: 82 331 290 0307

Email: myeun@sun20.aslti.re.kr

Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji

University, Yongin, Korea. 449-728 bhnahebio@server.myongji.ac.kr

Seg primer: M13 Reverse Primer.

Location/Qualifiers

1. 247

/organism="Oryza sativa"

/cultivar="Milyang23"

/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:

XhoI; Directional cDNA library inserted into lambda ZAPII

vector at 5' end with EcoRI and 3' end with Xho I site."

/db_xref="taxon:4530"

/clone="97SN1784"

/clone_1lb="Rice Immature Seed Lambda ZAPII cDNA Library"

/tissue_type="Immature Seed"

/dev_stage="5 days after pollination"

/lab_host="E. coli SOLR"

BASE COUNT 7 a 16 c 21 g 34 t 169 others

ORIGIN

Query Match 19.5%; Score 25; DB 12; Length 247;

Best Local Similarity 11.6%; Pred. No. 1.24e-05;

Matches 14; Conservative 60; Mismatches 45; Indels 2; Gaps 2;

Db 102 VHSNMNRCNSVYVBYATCDYBHYBDRANHVDTICTNDRCYCNVTASDNGTSATKR 161

Cp 125 GCCCAGCAAGATATCTTAT-TCTTTAAAGTTGAGATGACACTTGGAGATCGA 67

	source	1. .300 /organism="Caenorhabditis elegans" /strain="CB1489 him-8(e1489)" /db_xref="taxon:6239" /clone="yk429g" /clone_lib="Yuji Kohara unpublished cDNA library" /dev_stage="varied" /sex="hermaphrodite, male" /tissue_type="whole animal"
BASE COUNT	138 a 27 c 76 g 59 t	
ORIGIN		
Query Match	18.8%; Score 24; DB 9; Length 300;	
Best Local Similarity	83.3%; Pred. No. 1,29e-04;	
Matches	30; Conservative 0; Mismatches 6; Indels 0; Gaps 0;	
Db	72 TGATAGCTGAAGCGGTAGACGAAAAAATGATCC 107 	
Cp	69 TGAAGA GTTAAGAGGGGTAGACAAAATAATGACC 34	
RESULT	12	EST 29-AUG-1997
LOCUS	C53653 300 bp mRNA	
DEFINITION	C.elegans cDNA clone yk322e5 : 3' end, single read, mRNA sequence.	
ACCESSION	C53653	
NID	92391410	
KEYWORDS	EST: EST(expressed sequence tag).	
SOURCE	Caenorhabditis elegans (strain:CB1489 him-8(e1489)) varied hermaphrodite, male whole animal cDNA to mRNA, clone_lib:Yuji kohara unpublished cDNA library clone:yk322e5.	
ORGANISM	Caenorhabditis elegans	
REFERENCE	Eukaryotae: Metazoa: Nematoda, Secernentea, Rhabditiida: Rhabditiinae: Rhabditiidae: Pelodierinae: Caenorhabditis 1 (sites)	
AUTHORS	Kohara,Y., Motohashi,T., Tabata,H., Shin-i,T., Watanabe,H., Sano,M., Miyata,A., Ohba,T., Mtani,I.Y., Uesugi,H., Sugizura,I., Obana,M., Sumimoto,A., Iida,K. and Nishigaki,A.	
TITLE	Expression map of the C.elegans genome	
JOURNAL	Unpublished (1997)	
REFERENCE	2 (bases 1 to 300)	
AUTHORS	Kohara,Y.	
TITLE	Direct Submission	
JOURNAL	Submitted (23-JUL-1997) to the DDBJ/EMBL/GenBank databases. Yuji Kohara, National Institute of Genetics, Gene Network Lab; Yata 1111, Mishima, Shizuoka 411, Japan (E-mail:ykohara@elab.nig.ac.jp, Tel:+81-559-81-6854, Fax:+81-559-81-6855)	
FEATURES	Location/Qualifiers	
source	1..300 /organism="Caenorhabditis elegans" /strain="CB1489 him-8(e1489)" /db_xref="taxon:6239" /clone="yk322e5" /clone_lib="Yuji Kohara unpublished cDNA library" /dev_stage="varied" /sex="hermaphrodite, male" /tissue_type="whole animal"	
BASE COUNT	134 a 27 c 73 g 62 t 4 others	
ORIGIN		
Query Match	18.8%; Score 24; DB 9; Length 300;	
Best Local Similarity	83.3%; Pred. No. 1,29e-04;	
Matches	30; Conservative 0; Mismatches 6; Indels 0; Gaps 0;	
Db	74 TGATAAGTGAAGCGGTAGACGAAAAAATGATC 109 	
Cp	69 TGAAGA GTTAAGAGGGGTAGACAAAATAATGACC 34	
RESULT	13	EST 29-AUG-1997
LOCUS	C55340 300 bp mRNA	
DEFINITION	C.elegans cDNA clone yk202f6 : 3' end, single read, mRNA sequence.	
ACCESSION	C55340	
NID	92399941	

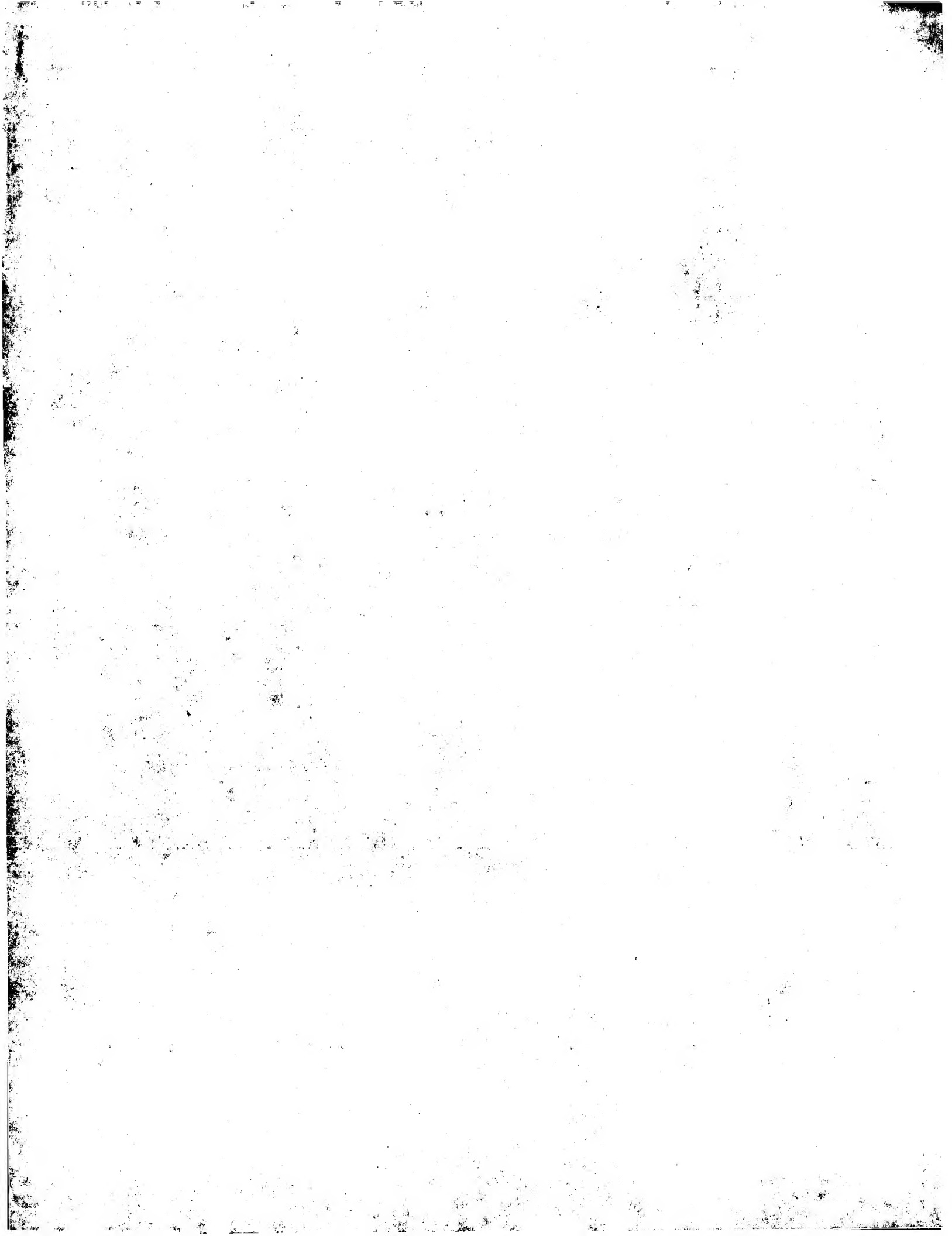
```

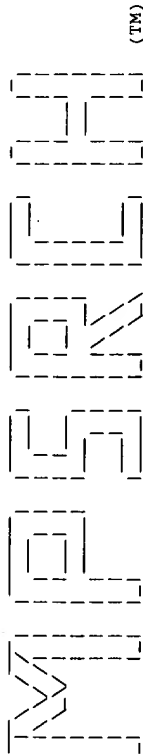
KEYWORDS      EST: EST(expressed sequence tag);
SOURCE        Caenorhabditis elegans (strain:CB1489 him-8(e1489)) varied
              hermaphrodite, male whole animal cDNA to mRNA, clone_1lb:Yuji
              Kohara unpublished cDNA library clone:YK202f6.
ORGANISM      Caenorhabditis elegans
              Eukaryote; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
              Rhabditina; Rhabditioidea; Rhabditidae; Peloderiinae; Caenorhabditis
              1 (siles)
REFERENCE      Kohara,Y., Motohashi,T., Tabara,H., Shin-I,T., Matanabe,H.,
AUTHORS       Sano,M., Miyata,A., Ohba,T., Mitani,Y., Uesugi,H., Sugita,I.,
              Obara,M., Sugimoto,A., Iida,K. and Nishigaki,A.
JOURNAL       Expression map of the C.elegans genome
              Unpublished (1997)
              2 (bases 1 to 300)
TITLE         Kohara,Y.
REFERENCE     Direct Submission
AUTHORS       Submitted (22-JUL-1997) to the DDBJ/EMBL/GenBank databases. Yuji
JOURNAL       Kohara, National Institute of Genetics, Gene Network Lab, Yata
              111, Mishima, Shizuoka 411, Japan (E-mail:Ykohara@lab.nig.ac.jp,
              Tel:+81-559-81-6854, Fax:+81-559-81-6855)
FEATURES
  source
    1..300
    /organism="Caenorhabditis elegans"
    /strain="CB1489 him-8(e1489)"
    /db_xref="taxon:6239"
    /clone="YK202f6"
    /clone_1lb="Yuji Kohara unpublished cDNA library"
    /dev_stage="varied"
    /sex="hermaphrodite, male"
    /tissue_type="whole animal"
BASE COUNT   138 a                27 c                76 g                58 t                1 others
ORIGIN
Query Match      18.8%; Score 24; DB 9; Length 300;
Best Local Similarity 83.3%; Pred. No. 1.29e-04;
Matches          30; Conservative 0; Mismatches 6; Indels 0; Caps 0;

Db      72 TCGATAGTGTAAAGCGTGAGAAAAAATGATCC 107
        ||| ||||| ||| |||| ||||| ||| |||
CP      69 TCGAAAGTGTAAAGCGTGAGCAAAAAAATGACCC 34

RESULT  14
LOCUS   CELK034GYR      349 bp      mRNA      EST      06-JAN-1997
DEFINITION C.elegans cDNA clone YK34g11 : 3' end, single read, mRNA sequence.
ACCESSION D33694
          G524621
NID
KEYWORDS EST: EST(expressed sequence tag);
SOURCE   Caenorhabditis elegans (strain:CB1489 him-8(e1489)) varied
          Hermaphrodite, male whole animal cDNA to mRNA, clone_1lb:Yuji
          Kohara unpublished cDNA.
ORGANISM Caenorhabditis elegans
          Eukaryote; mltochondrial eukaryotes; Metazoa; Nematoda;
          Secernentea; Rhabditia; Rhabditina; Rhabditidae;
          Rhabditioidea; Peloderiinae; Caenorhabditis.
          1 (bases 1 to 349)
REFERENCE Kohara,Y., MitsuKI,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
AUTHORS   Tabara,H.
TITLE     Toward an expression map of the C.elegans genome
JOURNAL   Unpublished (1994)
REFERENCE 2 (bases 1 to 349)
AUTHORS   Kohara,Y.
TITLE     Direct Submission
JOURNAL   Submitted (01-JUL-1994) to the DDBJ/EMBL/GenBank databases. Yuji
          Kohara, National Institute of Genetics, Gene Library Lab, Yata
          1111, Mishima, Shizuoka 411, Japan (E-mail:Ykohara@lab.nig.ac.jp,
          Tel:0559-81-6854, Fax:0559-81-6855)
FEATURES
  source
    1..349
    /organism="Caenorhabditis elegans"
    /strain="CB1489 him-8(e1489)"
    /db_xref="taxon:6239"
    /clone_1lb="Yuji Kohara unpublished cDNA"

```



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 05:09:04 1998; MasPar time 219.03 Seconds
Tabular output not generated. 981.093 Million cell updates/sec

Title: >US-08-530-112A-2058
Description: (1-120) from US08530112A.seq

Perfect Score: 119
N.A. Sequence: 1 GATCTTTACTCTTGAGAAAT.....AAAGTCCTNTAAAGCAA 120
Comp: CTAGAAATGAGAACTCTTTA.....TTTCGAGGNATTTTCGTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-est55

Database: 1:em_est1 2:em_gss1 3:em_gss2 4:em_gss3

Database: genbank-est107
5:gb_est1 6:gb_est10 7:gb_est11 8:gb_est12 9:gb_est13
10:gb_est14 11:gb_est15 12:gb_est16 13:gb_est17
14:gb_est18 15:gb_est19 16:gb_est2 17:gb_est20
18:gb_est21 19:gb_est3 20:gb_est4 21:gb_est5 22:gb_est6
23:gb_est7 24:gb_est8 25:gb_est9 26:gb_gss1 27:gb_gss2
28:gb_gss3 29:gb_gss4

Statistics: Mean 8.922; Variance 1.853; scale 4.815

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	119	100.0	404	8	AA479058	zvl8g02.s1 Soares NHM	3.15e-178
2	116	97.5	117	6	C21276	HUMGS0002342, Human Ge	2.03e-172
3	34	28.6	252	12	AA754459	97SN1787 Rice Immature	3.21e-23
4	32	26.9	247	12	AA754458	97SN1784 Rice Immature	3.92e-20
5	26	21.8	247	12	AA754458	97SN1787 Rice Immature	2.48e-11
6	25	21.0	252	12	AA754459	97SN1787 Rice Immature	6.01e-10
7	24	20.2	2275	11	AF034173	Homo sapiens ntcon2.co	1.37e-08
8	22	18.5	310	19	N57611	yz18f01.r1 Homo sapien	5.68e-06
9	21	17.6	322	23	AA080479	mn87c10.r1 Stratagene	1.03e-04
10	21	17.6	387	6	AA396585	mp23a12.r1 Life Tech m	1.03e-04
11	20	16.8	159	20	R57924	F6744 Fetal heart Homo	1.69e-03
12	20	16.8	286	5	T98031	ye30a02.s1 Homo sapien	1.69e-03
13	20	16.8	301	16	R68058	yh98e12.s1 Homo sapien	1.69e-03

C	14	20	16.8	306 25	AA347342	EST53624 Fetal heart I	1.69e-03
C	15	20	16.8	316 29	AQ249041	TL7P4-Sp6 TAMU Arabido	1.69e-03
C	16	20	16.8	377 7	HUM243C10B	Human aorta cDNA 5'-en	1.69e-03
C	17	20	16.8	392 18	AI057200	oz11a02.x1 Soares feta	1.69e-03
C	18	20	16.8	410 20	W56430	zd15g03.r1 Soares feta	1.69e-03
C	19	20	16.8	432 16	R63155	y101a11.s1 Homo sapien	1.69e-03
C	20	20	16.8	453 20	W56314	zd15g03.s1 Soares feta	1.69e-03
C	21	20	16.8	459 10	AA424226	zv54g08.r1 Soares test	1.69e-03
C	22	20	16.8	477 8	AA279262	zs84h09.r1 NCI_CGAP_GC	1.69e-03
C	23	20	16.8	480 22	AA024406	ze74e09.s1 Soares feta	1.69e-03
C	24	20	16.8	484 28	AQ186675	HS_3110.AL.F12.T7 CIT	1.69e-03
C	25	20	16.8	488 16	H27910	yo79e11.s1 Homo sapien	1.69e-03
C	26	20	16.8	513 26	B61705	TL8J21TR TAMU Arabidop	1.69e-03
C	27	20	16.8	519 28	AQ189629	HS_3219.BL.H10.T7 CIT	1.69e-03
C	28	20	16.8	585 18	AI181130	ub95b11.r1 Soares mous	1.69e-03
C	29	20	16.8	690 27	AQ057470	CIT-HSP-2340J10.TR CIT	1.69e-03
C	30	19	16.0	299 20	N94210	za27b07.r1 Homo sapien	2.52e-02
C	31	19	16.0	300 7	CELK126DYR	C.elegans cDNA clone y	2.52e-02
C	32	19	16.0	300 9	C54053	C.elegans cDNA clone y	2.52e-02
C	33	19	16.0	300 9	C59481	C.elegans cDNA clone y	2.52e-02
C	34	19	16.0	301 6	SSEET9	S.scrofa mRNA (clone E	2.52e-02
C	35	19	16.0	312 24	AA326266	EST29429 Cerebellum II	2.52e-02
C	36	19	16.0	316 19	N56191	J8473F Homo sapiens CD	2.52e-02
C	37	19	16.0	329 26	B97979	P23D22TR IGF Arabidop	2.52e-02
C	38	19	16.0	390 17	AU015448	Mus musculus 2-cell em	2.52e-02
C	39	19	16.0	421 10	AA422550	Vf14G05.s1 Knowles Sol	2.52e-02
C	40	19	16.0	422 26	FR0003452	F.rubripes GSS sequenc	2.52e-02
C	41	19	16.0	430 27	AQ063408	CIT-HSP-2346F24.TR CIT	2.52e-02
C	42	19	16.0	531 14	AA914535	vz02a10.r1 Soares mous	2.52e-02
C	43	19	16.0	621 27	AQ055791	CIT-HSP-2340O8.TF CIT-	2.52e-02
C	44	19	16.0	657 29	AQ231341	HS_2053.AL.E10_MR CIT	2.52e-02
C	45	19	16.0	661 27	B66364	CIT-HSP-2005B24.TFC CI	2.52e-02

ALIGNMENTS

RESULT 1
LOCUS AA479058 404 bp mRNA EST 08-AUG-1997
DEFINITION zvl8g02.s1 Soares NHMPu S1 Homo sapiens cDNA clone 754034 3', mRNA
ACCESSION AA479058
NID G2207614
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae;
Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Le.N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 279.
Location/Qualifiers
1. 404

FEATURES
source

/organism="Homo sapiens"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site.1: Not I;
Site.2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus

NbHPU, and fetal heart NbHH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as a tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

/db_xref="taxon:9606"

/clone="754034"

/clone_lib="Soares NhHPU S1"

/tissue_type="pooled human melanocyte, fetal heart, and

pregnant uterus"

/lab_host="DH10B"

complement(<1..>404)

/db_xref="GDB:5976908"

114 a 74 c 94 g 122 t

BASE COUNT
ORIGIN

Query Match 100.0%; Score 119; DB 8; Length 404;
Best Local Similarity 99.2%; Pred. No. 3.15e-178;
Matches 119; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 4 TTGCTTTTATAGGAGCTTTTCGGCGCTCCCATTTATTCCTTATACAGCTCAGTTTGT 63

||||| 117 |

Cp 120 TTGCTTTTANGAGCTTTTCGGCGCTCCCATTTATTCCTTATACAGCTCAGTTTGT 61

||||| 117 |

Db 64 CTAGTGATAGGACAGCTTTCTGAGAAAGCTTATTCATTCTCAAGATAAGATC 123

||||| 117 |

Cp 60 CTAGTGATAGGACAGCTTTCTGAGAAAGCTTATTCATTCTCAAGATAAGATC 1

||||| 117 |

RESULT 2 C21276 117 bp DNA EST 01-OCT-1996
LOCUS HUMGS0002342, Human Gene Signature, 3'-directed cDNA sequence, mRNA
DEFINITION
ACCESSION C21276
NID g1622386
KEYWORDS EST; Gene Signature; GS; EST(expressed sequence tag); BodyMap; gene expression.
SOURCE One or more human adult tissue.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Okubo,K.
TITLE BodyMap: human gene expression database
JOURNAL Unpublished (1996)
REFERENCE 2 (bases 1 to 117)
AUTHORS Okubo,K.
TITLE Direct Submission
JOURNAL Submitted (19-SEP-1996) to the DDBJ/EMBL/GenBank databases. Kousaku Okubo, Osaka University, Institute for Molecular and Cellular Bio; 1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan (E-mail:kousaku@imcb.osaka-u.ac.jp, Tel:06-877-5111(ex.3315), Fax:06-877-1922)

COMMENT We are not submitting the same cDNA sequence redundantly to DDBJ since 1993. For the abundance information of clones with this sequence in this library and as well as in other 3'-directed libraries, see, http://www.imcb.osaka-u.ac.jp/bodymap/. The sequences of the clones represented by this GS sequences is also found there.

FEATURES

Location/Qualifiers

1..117

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 42 a 24 c 22 g 28 t 1 others

ORIGIN

Query Match 97.5%; Score 116; DB 6; Length 117;

Best Local Similarity 100.0%; Pred. No. 2.03e-172;

Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GATCTTTACTTTCAGAAATGAATTAAGCTTTCTCTCAGAAATGCTGCTCCCTATACACTAG 60

||||| 117 |

Qy 1 GATCTTTACTTTCAGAAATGAATTAAGCTTTCTCTCAGAAATGCTGCTCCCTATACACTAG 60

||||| 117 |

Db 61 ACAAACGTAGCCTGTATAAGATAAATGGAGCGCCGAGAAAGCTCCNTAAAAAGC 117

||||| 117 |

Qy 61 ACAAACGTAGCCTGTATAAGATAAATGGAGCGCCGAGAAAGCTCCNTAAAAAGC 117

||||| 117 |

RESULT 3 AA754459 252 bp mRNA EST 20-JAN-1998

LOCUS 97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa

DEFINITION cDNA clone 97SN1787, mRNA sequence.

ACCESSION AA754459

NID g2801165

KEYWORDS EST.

SOURCE rice.

ORGANISM Oryza sativa

Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;

Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;

Poales; Poaceae; Oryza.

REFERENCE 1 (bases 1 to 252)

AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,

Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,

Lee,M.C. and Eun,M.Y.

Large-scale Sequencing Analysis of ESTs from Rice Immature Seed

Unpublished (1998)

CONTACT: Eun M.Y.

Department of Cytogenetics

National Inst. of Agri. Sci. and Tech, RDA

Suwon, Kyungdo, Korea

Tel: 82 331 290 0301

Fax: 82 331 290 0307

Email: myeungsun20.asti.re.kr

Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji

University, Yongin, Korea. 449-728 bnhahm@bioserver.myongji.ac.kr

Seq primer: M13 Reverse Primer.

Location/Qualifiers

1..252

/organism="Oryza sativa"

/cultivar="Milyang23"

/note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:

XhoI; Directional cDNA library inserted into lambda ZAPII

vector at 5' end with EcoRI and 3' end with Xho I site."

/db_xref="taxon:4530"

/clone="97SN1787"

/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"

/tissue_type="Immature Seed"

/dev_stage="5 days after pollination"

/lab_host="E. coli SOLR"

BASE COUNT 5 a 21 c 12 g 35 t 179 others

ORIGIN

Query Match 28.6%; Score 34; DB 12; Length 252;

Best Local Similarity 12.3%; Pred. No. 3.21e-23;

Matches 13; Conservative 57; Mismatches 35; Indels 1; Gaps 1;

Db 142 WBHYNTKCTASGHTSTNYDVKSTNTWGTBYSYDKSMHGYWCSBVKYHKVTTTATR 201

||||| 117 |

Cp 114 TTTTANGAGCTTTTCGGCGCTCCCATTTATTCCTTATACAGCTCAGTTTGTCTAGT 55

||||| 117 |

Db 202 SYTCVRKYCVMMTKVKVKKYHVVBGCHB-TDSCKCKTMMTNKHV 246

||||| 117 |

Cp 54 TATAGGACAGCATTTCTGAGAGAAAGCTTATTCATTCTCAAGAG 9

||||| 117 |

RESULT 4 AA754458 247 bp mRNA EST 20-JAN-1998

LOCUS 97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa

DEFINITION cDNA clone 97SN1784, mRNA sequence.

ACCESSION AA754458

NID g2801164

5

FEATURES

29 TACTCTAGAG

FEATURES

```

Oy 7 TACTCTGAGAAATGATAAGCTTCTCTCAG 38
||||| |||| | ||| |||||| |||||
RESULT 12
LOCUS T98031 286 bp mRNA EST
DEFINITION Ye30a02.s1 Homo sapiens cDNA clone 119210 3'.
ACCESSION T98031
NID 9747376
KEYWORDS EST.
SOURCE human clone=119210 library-Stratagene lung (#937210)
vector-pBluescript SK- host-SOLR cells (kanamycin resistant)
primer=21ml3 Rsite2-EcoRI Rsite2-XhoI Normal lung tissue from a 72
year old male. Cloned unidirectionally. Primer: Oligo dT. Average
insert size: 1.0 kb; Uni-ZAP XR vector; 5' adaptor sequence:
5'-CAATTCGGCAGGAG-3'; 3' adaptor sequence:
5'-CTCGAGTTTCTTTTCTTTTCTTTT-3'.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 286)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevasakis,E.,
Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
WashU-Merck EST Project
Unpublished (1995)
TITLE
JOURNAL
COMMENT
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 152
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES
source
Location/Qualifiers
1..286
/organism="Homo sapiens"
/clone="119210"
BASE COUNT 76 a 59 c 54 g 96 t 1 others
ORIGIN
Query Match 16.8%; Score 20; DB 5; Length 286;
Best Local Similarity 88.5%; Pred. No. 1.69e-03;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Db 123 CTATTAAATCTCTCAAGCAAGAT 148
||||| ||||| ||||| ||||| |||||
Cp 27 CTATTCAATCTCTCAAGATAAGAT 2
||||| ||||| ||||| ||||| |||||
RESULT 13
LOCUS R68058 301 bp mRNA EST
DEFINITION Yh98e12.s1 Homo sapiens cDNA clone 137806 3'.
ACCESSION R68058
NID 9841575
KEYWORDS EST.
SOURCE human clone=137806 library-Soares placenta NB2HP vector=pt7T3D
(Pharmacia) with a modified polylinker host=DH10B (ampicillin
resistant) primer=SP6 Rsite1-Not I Rsite2-Eco RI Female placenta
obtained at birth (full term). 1st strand cDNA was primed with a
Not I - oligo(dT) primer [5',
AAGTCGAGATCCGCGCCGAGCAATTTTCTTTTCTTTTCTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the modified pT73
vector. Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bernaldo.
Homo sapiens
Eucaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;

```

```

REFERENCE
AUTHORS
1 (bases 1 to 301)
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archontes; Primates; Catarrhini; Hominidae; Homo.
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
TITLE
JOURNAL
COMMENT
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 152
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES
source
Location/Qualifiers
1..301
/organism="Homo sapiens"
/clone="137806"
BASE COUNT 72 a 68 c 55 g 100 t 6 others
ORIGIN
Query Match 16.8%; Score 20; DB 16; Length 301;
Best Local Similarity 82.8%; Pred. No. 1.69e-03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db 273 CTCTCTNAGAAATCTGTCCCATCCA 301
||||| ||||| ||||| ||||| |||||
Qy 28 CTCTCTCAGAAATGCTGCTCTATACA 56
||||| ||||| ||||| ||||| |||||
RESULT 14
LOCUS AA347342 306 bp mRNA EST
DEFINITION EST53624 Fetal heart II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA347342
NID g1995979
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eucaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
Homo.
REFERENCE 1 (bases 1 to 306)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Whi,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitchugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,X.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
JOURNAL
MEDLINE
95026280
COMMENT Other_ESTs: EST53623

```

Contact: Kerlavage, AR
Bioinformatics

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423

Email: arke@lavetigr.org For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/cdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES
SOURCE

Location/Qualifiers
1. .306

```

/organism="Homo sapiens"
/note="Organ: heart; Vector: pBluescript SK-; Site_1:
EnoB1: site_2, ybat"

```

```
/db_xref="ATCC (inhost):148609"
/db_xref="taxon:9606"

```

```
/dev_stage="fetus"
<1. .>306
```

ORIGIN	BASE COUNT	MRNA
105 a	46 c	<1. .>306
		53 g
		102 t

Query Match	16.88;	Score 20;	DB 25;	Length 306;
Best Local Similarity	73.88;	Pred. No. 1.69e-03;		
Matches	31;	Conservative	0;	Mismatches 11; Indels 0; Gaps 0;

Db 55 ACAGGTCATTGGCCCTACGTACTGTACAACATTTACTCGA 96
 ||||| ||| | ||||| ||| | ||||| ||| |
Cp 76 ACAGGCTCAGTTTTGCTAGTGTATAGGACAGCATTTTCGA 359

RESULT	15
LOCUS	AQ249041 316 bp DNA GSS 05-OCT-1998
DEFINITION	T1P4-Sp6 TAMU Arabidopsis thaliana genomic clone T1P4, genomic survey sequence.

ACCESSION	AQ249041
NID	g3699124

KEYWORDS GSS
SOURCE tha

ORGANISM *Arabidopsis thaliana*

Embryophyta: Viridiplantae: Charophyta/Embryophyta group;
Embryophyta: Tracheophyta, euphyllophytes; Spermatophyta;
Magnoliophyta: eudicotyledons; Rosidae; Caprales; Brassicaceae;
Arabidopsis.

REFERENCE	1 (bases 1 to 316)
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100
101	101
102	102
103	103
104	104
105	105
106	106
107	107
108	108
109	109
110	110
111	111
112	112
113	113
114	114
115	115
116	116
117	117
118	118
119	119
120	120
121	121
122	122
123	123
124	124
125	125
126	126
127	127
128	128
129	129
130	130
131	131
132	132
133	133
134	134
135	135
136	136
137	137
138	138
139	139
140	140
141	141
142	142
143	143
144	144
145	145
146	146
147	147
148	148
149	149
150	150
151	151
152	152
153	153
154	154
155	155
156	156
157	157
158	158
159	159
160	160
161	161
162	162
163	163
164	164
165	165
166	166
167	167
168	168
169	169
170	170
171	17

Ecker, J.

JOURNAL
Unpublished (1997)

COMMENT

Contact: Ecker J.

Aradiopsis Indiana genome center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104

Tel: 215-898-9384
Fax: 215-898-8780

Email: jecker@atgenome.bio.upenn.edu

Class: BAC ends

FEATURES	Location/Qualifiers
source	1. .316

```
/organism="Arabidopsis thaliana"
/strain="Columbia"
```

```

/note="Vector: Belobacii, Site_1: HindIII, site_2
HindIII; Produced by Rod Wing"

```

```
/db_xref="taxon:3702"  
/clone="T17P4"
```

```
/clone_lib="TAMU"  
/sex="hermaphrodite"
```

BASE COUNT	81 a	58 c	55 g	120 t	2 others
------------	------	------	------	-------	----------

ORIGIN

Query Match	16.8%;	Score 20;	DB 29;	Length 316
-------------	--------	-----------	--------	------------

Best Local Similarity	12.13;	Pred.NO. 1.09E-03;
Matches	32; Conservative	0; Mismatches 12; Indels 0; Gaps 0

```

Db 66 CTTTCTCGGCTATTGATTCCTATTCCAGACTATGATTGT 109
      |||||  |||  |||  |||||  |||  |||  |||  |||||
Cp 104 CTTTTCGGGCTCCATTATTTCCTATACAGGCTCAGTTTGT 61

```

Search completed: Tue Dec 22 05:12:50 1998
Job time : 226 secs.




 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution Rights by Oxford Molecular Ltd

Mpsrch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 05:13:09 1998; Maspar time 33.85 Seconds
 482.262 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-2058
 Description: (1-120) from US08530112A.seq
 Perfect Score: 119
 N.A. Sequence: 1 GATCTTACTCTTGAGAAAT.....AAAGCTCCNTAAAGCAAA 120
 Comp: CTAGAAATGAGAACTCTTTA.....TTTCGAGGNAITTTTCGTTT

Scoring table: TABLE default
 Gap 6

Mmatch STD : Dbase 0: Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database:

n-geneseq22
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39 40:part40

Statistics: Mean 6.952; Variance 4.158; scale 1.672

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description	Pred. No.
1	119	100	0	Human gene signature	2.55e-64
2	118	99	2	hML genomic DNA	1.24e-63
3	70	58	8	Thrombopoietin coding	3.66e-31
4	41	34	5	hML cDNA	1.50e-12
5	41	34	5	Human thrombopoietin	1.50e-12
6	36	30	3	Base substituted E.co	6.56e-09
7	35	29	4	Oligonucleotide probe	1.01e-07
8	33	27	7	Oligonucleotide probe	3.66e-03
9	25	21	0	Human GS HUMG02342-b	3.66e-03
10	25	21	0	Human mpl-1ligand gene	3.66e-03
11	25	21	0	Human mpl-1ligand gene	3.66e-03
12	25	21	0	Base substituted E.co	3.66e-03
13	24	20	2	Generic DNA sequence	1.28e-02

Result No.	Score	Query Match Length	ID	Description	Pred. No.
14	23	19	3	Generic DNA sequence	4.40e-02
15	22	18	5	Generic DNA sequence	1.48e-01
16	22	18	5	Generic DNA sequence	1.48e-01
17	22	18	5	Human endothelin-1 an	1.48e-01
18	22	18	5	Sequence encoding new	1.48e-01
19	21	17	6	Generic DNA sequence	4.90e-01
20	20	16	8	DC43 TSAR library gen	1.59e+00
21	20	16	8	DC43 TSAR library gen	1.59e+00
22	19	16	0	TSAR-9 library genera	5.00e+00
23	19	16	0	TSAR-9 library genera	5.00e+00
24	19	16	0	DC43 TSAR library gen	5.00e+00
25	19	16	0	DC43 TSAR library gen	5.00e+00
26	19	16	0	Generic DNA sequence	5.00e+00
27	19	16	0	Generic DNA sequence	5.00e+00
28	19	16	0	Sequence encoding new	5.00e+00
29	19	16	0	Fragment of human OCI	5.00e+00
30	18	15	1	TSAR-9 library genera	1.54e+01
31	18	15	1	TSAR-9 library genera	1.54e+01
32	18	15	1	Human gene signature	1.54e+01
33	18	15	1	Human IL-8 receptor-a	1.54e+01
34	18	15	1	Sequence encoding new	1.54e+01
35	18	15	1	Sequence encoding new	1.54e+01
36	18	15	1	Sequence encoding new	1.54e+01
37	18	15	1	Sequence encoding new	1.54e+01
38	18	15	1	Antiser specific cDNA	1.54e+01
39	18	15	1	NK cell antigen DX1 9	1.54e+01
40	18	15	1	Human CII-3 cDNA	1.54e+01
41	18	15	1	Human MKK3-interactin	1.54e+01
42	18	15	1	Chicken liver alpha-N	1.54e+01
43	18	15	1	Chicken liver alpha-N	1.54e+01
44	18	15	1	Chicken alpha-N-acety	1.54e+01
45	18	15	1	DNA sequence of plasm	1.54e+01

ALIGNMENTS

RESULT 1
 ID T21058 standard; cDNA to mRNA: 120 BP.
 AC T21058;
 DT 03-JUL-1996 (first entry)
 DE Human gene signature HUMG02342.
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 human; cloning; mapping; non-biased library; diagnosis; detection;
 cell typing; abnormal cell function; ss.
 OS Homo sapiens.
 PN W09514772-A1.
 PD 01-JUN-1995.
 PF 11-NOV-1994; J01916
 PR 12-NOV-1993; J01916
 PA (MATSU) MATSUBARA K.
 PI (MATSU) MATSUBARA K.
 PI (OKUBO) OKUBO K.
 DR WPI: 95-206931/27.
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 for diagnosis of abnormal cell function, by preparing cDNA that
 reflects relative abundance of corresp. mRNA in specific human
 tissues
 PS Claim 1: Page 791; 2245pp; Japanese.
 CC A single-stranded DNA (or its complementary strand or the corresp.
 double-stranded DNA) which comprises one of the 7837 "GS" sequences
 given in T19001-T26837 and which is able to hybridise to part of
 human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 sequences were obtained from 3'-directed cDNA libraries prepared
 from various human tissues; synthesis of cDNA was initiated from the
 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 untranslated sequence is unique to a particular mRNA species, almost
 all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 is constructed so as to reflect accurately the relative abundance of
 different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 determined (esp. using primers and probes derived from the GS
 sequences) as a means of diagnosing abnormal cell function or for
 recognising different cell types.

```

SQ Sequence 120 BP; 45 A; 24 C; 22 G; 28 T;
Query Match 100.0%; Score 119; DB 18; Length 120;
Best Local Similarity 100.0%; Pred. No. 2,55e-64;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 gatcttactctgagaagaataagcttctctcagaatcgtcctatacactag 60
   |||||||
QY 1 GATCTTACTCTTGAGAAAGATAAGCTTCTCTCAGAAATGCTGCTATACACTAG 60
   |||||||

Db 61 acaaaactgagcctgataagaataatgagcgccgaaagctccnlaaaagcaaa 120
   |||||||
QY 61 ACAAACTGAGCCTGTATAAGATAATGCGAGCCGCAAAAGCTCCNTAAAAAGCAAA 120

RESULT 2
ID Q94109 standard; DNA; 7849 BP.
AC Q94109;
DT 22-FEB-1996 (first entry)
DE hML genomic DNA.
KW Human; thrombopoietin; TPO; mpl ligand; hML; fragment polypeptide;
   megakaryocytopoietic cytokine receptor; thrombopoietic signal;
   EPO-domain fragment; erythropoietin; hEPO; haematopoietic cell;
   megakaryocyte; thrombocytopenia; myeloproliferative disease;
   inflammatory thrombocytosis; iron deficiency; EPO; platelet;
   red blood cell; progenitor; hML-2; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT prim_transcript 1166..7289
   /tag= a
FT exon 1161..1232
   /tag= b
FT intron 1233..2900
   /number= 1
FT exon 2901..3058
   /tag= c
   /tag= d
   /number= 2
FT signal_peptide 3046..3058
   /tag= e
   /note= "Encodes partial signal peptide"
FT intron 3059..33289
   /tag= f
FT exon 3290..3417
   /tag= g
   /number= 3
FT signal_peptide 3290..3339
   /tag= h
   /note= "Encodes partial signal peptide"
FT intron 3418..3703
   /tag= i
FT exon 3704..3790
   /tag= j
   /number= 4
FT intron 3791..5722
   /tag= k
FT exon 5723..5890
   /tag= l
   /number= 5
FT intron 5891..6126
   /tag= m
FT intron 5891..6138
   /tag= n
   /note= "Utilises alternative splice site, causes
   formation of cDNA encoding hML-2 comprising
   a four amino acid deletion"
FT exon 6127..7289
   /tag= o
   /number= 6
FT misc_signal 6790..6792
   /tag= p
FT misc_feature 7290
   /note= "Stop codon"

```

```

FT FT /tag= g
PN PN /standard_name= polyA_site
DB GB2285446-A.
PD 12-JUL-1995.
PF 21-DEC-1994; 025831.
PR 03-JAN-1994; US-176553.
PR 21-JAN-1994; US-185607.
PR 15-FEB-1994; US-196689.
PR 04-APR-1994; US-223253.
PR 25-MAY-1994; US-249376.
PR 02-DEC-1994; US-348658.
PR 02-DEC-1994; US-348657.
PA (GENETH ) GENENTECH INC.
PI Baton DL, de Sauvage FT.
DR WPI: 95-234018/31.
DR P-PSDB; R76164.
PT Thrombopoietin polypeptide, ligand for mpl cytokine receptor -
   useful for treating thrombocytopenia and related diseases
PS Example 8; Fig 14; 192pp; English.
CC This sequence represents a genomic clone encoding human thrombopoietin
   (TPO), also known as mpl ligand (hML). The human gene comprises 6 exons
   within 7 kb of genomic DNA. Exons 1 and 2 contain 5' untranslated
   sequence and the initial four amino acids of the signal peptide. The
   remainder of the secretory signal and the first 26 amino acids of the
   mature protein are encoded within exon 3. The entire carboxyl domain
   and 3' untranslated sequence, as well as approx. 50 amino acids of the
   erythropoietin-like domain are encoded within exon 6. The four amino
   acids involved in the deletion observed in hML-2 (hTPO-2) are encoded at
   the 5' end of exon 6. Analysis of human genomic DNA by Southern blot
   indicated the gene for TPO is present in a single copy. The chromosomal
   location of the gene was determined by fluorescent in situ hybridisation
   CC which mapped to chromosome 3q27-28. The protein encoded by this sequence
   CC binds to mpl which is a megakaryocytopoietic cytokine receptor which may
   CC transduce a thrombopoietic signal. The N-terminal fragment of hML shows
   CC high levels of homology with human erythropoietin (hEPO). This
   CC constitutes the EPO-domain. The EPO-domain spans from Cys7 to Cys151
   of mature hML.
SQ Sequence 7849 BP; 1985 A; 2151 C; 1881 G; 1831 T;

Query Match 99.2%; Score 118; DB 16; Length 7849;
Best Local Similarity 99.2%; Pred. No. 1.24e-63;
Matches 118; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 7252 gatcttactctgagaagaataagcttctctcagaatcgtcctatacactag 7311
   |||||||
QY 1 GATCTTACTCTTGAGAAAGATAAGCTTCTCTCAGAAATGCTGCTATACACTAG 60
   |||||||

Db 7312 acaaaactgagcctgataagaataatgagcgccgaaagctccnlaaaagcaaa 7370
   |||||||
QY 61 ACAAACTGAGCCTGTATAAGATAATGCGAGCCGCAAAAGCTCCNTAAAAAGCAAA 119
   |||||||

RESULT 3
ID Q99554 standard; cDNA to mRNA; 1721 BP.
AC Q99554;
DT 29-APR-1996 (first entry)
DE Thrombopoietin coding sequence.
KW Thrombopoietin; TPO; humoral factor; platelet; antibody; therapy; AIDS;
   KW thrombocytopenia; hypoplastic anaemia; thrombotic thrombocytopenia;
   KW disseminated intravascular coagulation syndrome; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT msc_feature 78..1344
   /tag= a
   /note= "see Q99553 (claimed)"
FT cds 25..1086
   /tag= b
   /product= thrombopoietin
FT signal_peptide 25..87
   /tag= c
FT mat_peptide 88..1083
   /tag= d
PN W09521919-A2.

```


PD 17-AUG-1995. J00208.
 PF 14-FEB-1995: JP-039090.
 PR 14-FEB-1994: US-212154.
 PR 14-MAR-1994: JP-079842.
 PR 25-MAR-1994: US-221020.
 PR 01-APR-1994: JP-155126.
 PR 01-JUN-1994: JP-167328.
 PR 15-JUN-1994: US-278083.
 PR 20-JUL-1994: JP-193159.
 PR 17-AUG-1994: JP-227159.
 PR 17-AUG-1994: JP-193916.
 PR 18-AUG-1994: US-320300.
 PR 11-OCT-1994: JP-304157.
 PR 01-NOV-1994: JP-298659.
 PR 01-DEC-1994: US-361811.
 PR 22-DEC-1994: JP-341200.
 PR 28-DEC-1994: US-381478.
 PR 31-JAN-1995: US-381478.
 PA (KIRI) KIRIN BREWERY KK.
 PI Akahori H, Iwamatsu A, Kato T, Kuroki R, Miyazaki H;
 PI Muto T, Ohgami K, Shimizu T;
 DR WPI: 95-293120/38.
 DR P-PSDB: R81341.
 PT Thrombopoietin polypeptide which specifically stimulates or
 PT increases platelet production - useful in treatment of
 PT thrombocytopenia, also related DNA and vectors
 PS Example 22: Page 282-284, 383pp; English.
 CC This sequence represents the coding sequence for a thrombopoietin (TPO)
 CC protein. This sequence contains the sequence represented by 099553. TPO
 CC is a humoral factor capable of promoting platelet production. The
 CC encoded protein can be purified by using a gel affinity column where Mpl
 CC has been coupled to a resin. 099551-099553 also encode TPO proteins.
 CC The DNA sequences are inserted into vectors which are used to transform
 CC prokaryotic and eukaryotic host cells. Using the proteins encoded by
 CC these sequences, and derivatives of them, antibodies specifically
 CC immunoreactive with a TPO protein can be created. The antibodies, DNA
 CC sequences and vectors are used to isolate the protein sequences. TPO
 CC proteins can then be used in the treatment of platelet disorders. These
 CC include thrombocytopenia, hypoplastic anaemia, AIDS, disseminated
 CC intravascular coagulation syndrome and thrombotic thrombocytopenia.
 SQ Sequence 1721 BP; 413 A; 542 C; 371 G; 395 T;

Query Match 58.8%; Score 70; DB 17; Length 1721;
 Best Local Similarity 100.0%; Pred. No. 3.60e-31;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1623 gatcttactcttgagaataagatgcttctcagaatactgctctatacactag 1682
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1 GATCTTACTCTTGAGAAATGATAGCTTCTCTCAGAAATGCTCTATACACTAG 60

Db 1683 acaaaactga 1692
 ||||||||
 QY 61 ACAAACTGA 70

RESULT 4
 ID 094107 standard: cDNA: 1795 BP.
 AC 094107;
 DT 22-FEB-1996 (first entry)
 DE hml cDNA.
 KW Human: thrombopoietin; TPO: mpl ligand; hml: fragment polypeptide;
 KW megakaryocytopoietic cytokine receptor; thrombopoietic signal;
 KW EPO-domain fragment; erythropoietin; hEPO: haematopoietic cell;
 KW megakaryocyte; thrombocytopenia; myeloproliferative disease;
 KW inflammatory thrombocytosis; iron deficiency; EPO; platelet;
 KW red blood cell; progenitor; ss.
 OS Homo sapiens.
 FH key Location/Qualifiers
 FT 5'utr 1..215
 FT cds /*tag= a
 FT 216..1277
 FT /*tag= b
 FT /product= human mpl ligand

FT signal_peptide 216..278
 FT /*tag= c
 FT mat_peptide 279..274
 FT /*tag= d
 FT 3'utr 278..1795
 FT /*tag= e
 FT exon 229..356
 FT /*tag= f
 FT /note= "Presumed exon 3"

PN GB2285446-A.
 PD 12-JUL-1995.
 PF 21-DEC-1994: 025831.
 PR 03-JAN-1994: US-176553.
 PR 21-FEB-1994: US-185607.
 PR 15-FEB-1994: US-196689.
 PR 04-APR-1994: US-223263.
 PR 25-MAY-1994: US-249376.
 PR 02-DEC-1994: US-348658.
 PR 02-DEC-1994: US-348657.
 PA (GERTH) GENENTECH INC.
 PI Baton DL, de Sauvage FU;
 PI WPI: 95-234018/31.
 DR P-PSDB: R76164.
 PT Thrombopoietin polypeptide, ligand for mpl cytokine receptor -
 PT useful for treating thrombocytopenia and related diseases
 PS Claim 10; Fig 1; 192pp; English.
 CC This sequence encodes human thrombopoietin (TPO), also known as mpl
 CC ligand (hml). The protein encoded by this sequence binds to mpl which
 CC is a megakaryocytopoietic cytokine receptor which may transduce a
 CC thrombopoietic signal. The N-terminal fragment of hml shows high levels
 CC of homology with human erythropoietin (hEPO). This constitutes the EPO-
 CC domain. The EPO-domain spans from Cys7 to Cys151 of mature hml. The
 CC full length hml protein and fragment polypeptides derived from the
 CC EPO-domain, may be used to stimulate survival, proliferation,
 CC differentiation, and maturation of haematopoietic cells, esp.
 CC megakaryocytes. They are useful in the treatment and prevention
 CC of thrombocytopenia, myeloproliferative diseases, inflammatory
 CC thrombocytosis, iron deficiency etc. Chimera's of the fragment
 CC polypeptides and EPO can regulate growth and proliferation of both
 CC platelet and red blood cell progenitors.
 SQ Sequence 1795 BP; 420 A; 586 C; 382 G; 407 T;

Query Match 34.5%; Score 41; DB 16; Length 1795;
 Best Local Similarity 100.0%; Pred. No. 1.30e-12;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1737 gatcttactcttgagaataagatgcttctcagaata 1777
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1 GATCTTACTCTTGAGAAATGATAGCTTCTCTCAGAAA 41

RESULT 5
 ID T64318 standard: cDNA: 1795 BP.
 AC T64318;
 DT 21-MAY-1997 (first entry)
 DE Human thrombopoietin cDNA.
 KW TPO: thrombopoietin; C-terminal fragment; platelet count; increase;
 KW thrombocytopenia; acquired immune deficiency syndrome; AIDS; HIV;
 KW human immunodeficiency virus; cytokine; ds.
 OS Homo sapiens.
 FH key Location/Qualifiers
 FT cds 216..1277
 FT /*tag= a
 FT /product= thrombopoietin
 FT /note= "C-terminal domain is used for increasing
 FT platelet count"
 FT signal_peptide 216..278
 FT /*tag= b
 FT mat_peptide 279..1274
 FT /*tag= c
 FT US5593666-A.
 PD 14-JAN-1997.
 PF 16-AUG-1994: 291376.


```

RESULT 9
ID T26842 standard; cDNA to mRNA; 25 BP.
AC T26842;
DE 19-NOV-1996 (first entry)
KW Human GS HUMGS02342-based reverse transcriptase primer PA-6.
KW Gene signature: messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; primer; PCR; amplification;
KW polymerase chain reaction; yeast; ribosome; ss.
OS Synthetic.
PN MO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Example 10: Page 2186; 2245pp; Japanese.
CC This is the sequence of a primer based on the sequence of the human gene
CC signature (GS) sequence HUMGS02342 (T21342) which was used to reverse
CC transcribe mRNA isolated from human foetal liver polyA+ RNA for cloning
CC the cDNA encoding the human mpl-1 ligand, using the 5'-amplifinder (RM)
CC kit. The resultant cDNA strand was ligated to the amplifinder Anchor
CC primer (T26843) and the clone was amplified using PCR primers T26844 and
CC T26846.
CC The GS sequences (T19001-T26837) were obtained from 3'-directed cDNA
CC libraries prepared from various human tissues; synthesis of cDNA was
CC initiated from the 3'-end of mRNA by using poly(T) as the sole primer.
CC Each library is constructed so as to reflect accurately the relative
CC abundance of different mRNAs in the particular tissue from which it was
CC derived. The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS sequences)
CC as a means of diagnosing abnormal cell function or for recognising
CC different cell types.
CC Sequence 25 BP; 2 A; 8 C; 3 G; 12 T;
SO

Query Match 21.0%; Score 25; DB 23; Length 25;
Best Local Similarity 100.0%; Pred. No. 3.66e-03;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 ttctcgagcgtccattatcttcctt 25
|||||
Cp 103 TTTTCGGCGCTCCCATTTTCTTCTT 79

RESULT 10
ID T26845 standard; cDNA to mRNA; 37 BP.
AC T26845;
DE 19-NOV-1996 (first entry)
KW Human mpl-1 ligand gene PA-4 PCR primer.
KW Gene signature: messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; primer; PCR; amplification;
KW polymerase chain reaction; ss.
OS Synthetic.
PN MO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues

```

```

PS Example 10: Page 2187; 2245pp; Japanese.
CC Primers T26844-6 were used to PCR amplify the gene sequence encoding the
CC human mpl-1 ligand. T26844 and T26846 were used in a primary PCR whereas
CC T26845, which is internal to T26844, was used in a secondary nested
CC amplification reaction. The template for the PCR was cDNA derived human
CC foetal liver polyA+ RNA which was reverse transcribed using a primer
CC based on the human gene signature (GS) sequence HUMGS02342 (T21342).
CC The GS sequences (T19001-T26837) were obtained from 3'-directed cDNA
CC libraries prepared from various human tissues; synthesis of cDNA was
CC initiated from the 3'-end of mRNA by using poly(T) as the sole primer.
CC Each library is constructed so as to reflect accurately the relative
CC abundance of different mRNAs in the particular tissue from which it was
CC derived. The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS sequences)
CC as a means of diagnosing abnormal cell function or for recognising
CC different cell types.
CC Sequence 37 BP; 8 A; 10 C; 10 G; 9 T;
SO

Query Match 21.0%; Score 25; DB 23; Length 37;
Best Local Similarity 100.0%; Pred. No. 3.66e-03;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 13 tgcataagcagcagcattctgagag 37
|||||
Cp 56 TGTATAGGACACGACATTTCTGACAG 32

RESULT 11
ID T26844 standard; cDNA to mRNA; 37 BP.
AC T26844;
DE 19-NOV-1996 (first entry)
KW Human mpl-1 ligand gene PA-5 PCR primer.
KW Gene signature: messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; primer; PCR; amplification;
KW polymerase chain reaction; ss.
OS Synthetic.
PN MO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Example 10: Page 2187; 2245pp; Japanese.
CC Primers T26844-6 were used to PCR amplify the gene sequence encoding the
CC human mpl-1 ligand. T26844 and T26846 were used in a primary PCR whereas
CC T26845, which is internal to T26844, was used in a secondary nested
CC amplification reaction. The template for the PCR was cDNA derived human
CC foetal liver polyA+ RNA which was reverse transcribed using a primer
CC based on the human gene signature (GS) sequence HUMGS02342 (T21342).
CC The GS sequences (T19001-T26837) were obtained from 3'-directed cDNA
CC libraries prepared from various human tissues; synthesis of cDNA was
CC initiated from the 3'-end of mRNA by using poly(T) as the sole primer.
CC Each library is constructed so as to reflect accurately the relative
CC abundance of different mRNAs in the particular tissue from which it was
CC derived. The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS sequences)
CC as a means of diagnosing abnormal cell function or for recognising
CC different cell types.
CC Sequence 37 BP; 5 A; 13 C; 6 G; 13 T;
SO

Query Match 21.0%; Score 25; DB 23; Length 37;
Best Local Similarity 100.0%; Pred. No. 3.66e-03;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 13 ttcctataagcagcagcattctgtct 37
|||||

```

```

Cp      83  TCCTATACAGCCTCAGTTTGTCT 59

RESULT 12
ID      N81164 standard; DNA: 204 BP.
AC      N81164:
DT      08-NOV-1990 (first entry)
DE      Base substituted E.coli beta-galactosidase alpha-fragment.
KW      E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS      Escherichia coli.
FH      key
FH      Location/Qualifiers
FT      misc_feature
          19..69
          /tag= a
          /function= multiple cloning site
          187..204
          /tag= b

FT      primer_bind
          187..204
          /tag= b

PN      EP-285123-A.
PD      05-MAY-1988.
PF      30-MAR-1988; 105163.
PR      03-APR-1987; US-034819.
PI      (SUSO) SUOMEN SOKERI OY.
PI      Lehtovaara P, Knowles J, Kolvula A, Bamford J, Reinikainen T;
PT      WPI: 88-279927/40.
PT      Introducing random point mutations into nucleic acids -
PT      by prep of single stranded template, annealing a primer, elongation,
PT      pmiscorporation, completion of molecules and screening.
PS      Disclosure: P: English.
CC      Random point mutations were introduced into the alpha fragment of
CC      E.coli beta galactosidase. The wild type sequence was obtained as a
CC      single stranded template and an oligonucleotide was hybridised to
CC      it to generate a popn of DNA molecules which terminate at all
CC      possible nucleotide positions within a specified region. The
CC      variable 3' ends generated in this way are used as primers for
CC      reverse transcriptase. Nucleotides are misincorporated by the
CC      transcriptase and the molecules are completed to forms that can be
CC      amplified and then expressed in a suitable host-vector system.
CC      The sequence covers all 176 diff base substitutions, most of which
CC      occurred singularly in any given mutant.
CC      See also P80575.
SQ      Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 21.0%; Score 25; DB 1; Length 204;
Best Local Similarity 7.4%; Pred. No. 3,66e-03;
Matches 4; Conservative 33; Mismatches 16; Indels 1; Gaps 1;

Dd      132 hddhyvvybbbyvnyvnhnnccbnhvcnhvbnhnnhwayvrrdardvth 185
          ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Cp      104 CTTTCGCGCGCTCCATTATTCCTTATTCACAGCCTC-AGTTTGTCTAGTAT 52

RESULT 13
ID      Q70467 standard; DNA: 114 BP.
AC      Q70467:
DT      05-APR-1995 (first entry)
DE      Generic DNA sequence to generate a random TSAR peptide library.
DE      Generic DNA sequence to generate a random TSAR peptide library.
DE      TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW      effector domain; concatenated heterofunctional protein; linker;
KW      direct; rapid; detection; screening; treatment; generic; ss.
OS      Synthetic.
FH      key
FH      Location/Qualifiers
FT      misc_feature
          55..60
          /tag= a
          /note= "this sequence represents 'Z'; Z can be a
          sequence of 6, 9 or 12 nucleotides (see
          comments)"

FT      WO9418318-A.
PD      18-AUG-1994.
PF      01-FEB-1994; U00977.
PR      01-FEB-1993; US-013416.
PR      30-DEC-1993; US-176500.
PR      31-JAN-1994; US-189331.
PA      (UYNC-) UNIV NORTH CAROLINA.
PI      Fowlkes DM, Kay BK;

```

[illegible]

BASE COUNT	1727 a	1871 c	1616 g	1634 t	1 others
ORIGIN					
Query Match	99.2%	Score 118;	DB 25;	Length 6849;	
Best Local Similarity	99.2%	Pred. No. 5.03e-74;			
Matches 118;	Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;
Db	6252	GATCTTTACTCTTGGAATGAATAGCTTCTCAGAAATGCTGCTCCCTATACACTAG	6311		
Oy	1	GATCTTTACTCTTGGAATGAATAGCTTCTCAGAAATGCTGCTCCCTATACACTAG	60		
Db	6312	ACAAACTGAGCCTGTATAGGAATTAATGGAGCGCCGAAAAAGCTCCCTAAAAAGCAA	6370		
Oy	61	ACAAACTGAGCCTGTATAGGAATTAATGGAGCGCCGAAAAAGCTCCCTAAAAAGCAA	119		
LOCUS	2	HUMTA	7666 bp	DNA	PRI
DEFINITION	Human gene for thrombopoietin, exon1-exon6, complete cds.				
ACCESSION	D32046				
NID	9577319				
KEYWORDS	Thrombopoietin.				
SOURCE	Homo sapiens blood DNA, clone:lambdaHGT1.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 7666)				
TITLE	Sohma,T., Akahori,H., Seki,N., Horii,T., Ogami,K., Kato,T., Shimada,Y., Kawamura,K. and Miyazaki,H. Molecular cloning and chromosomal localization of the human thrombopoietin gene				
JOURNAL	FEBS Lett. 353 (1), 57-61 (1994)				
MEDLINE	95010765				
REFERENCE	2 (bases 1 to 7666)				
AUTHORS	Sohma,T.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-JUL-1994) to the DDBJ/EMBL/GenBank databases.				
MEDLINE	Yoshiaki Sohma, KIRIN Brewery Co., Ltd. Pharmaceutical Research Laboratory; 2-2, Soujamaichi Ichome, Maebashi, Gunma 371, Japan (Tel:0272-54-8618, Fax:0272-52-2307)				
FEATURES	Location/Qualifiers				
source	1. .7666				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="3"				
	/clone="lambdaHGT1"				
	/map="3q27"				
	/tissue_type="blood"				
	<1420. .1486				
exon	/number=1				
exon	3157. .3314				
exon	/number=2				
CDS	join(3302. .3314,3546. .3673,3960. .4046,5979. .6146,6383. .7048)				
	/codon_start=1				
	/product="thrombopoietin"				
	/db_xref="PID:d1007381"				
	/db_xref="PID:g577320"				
	/translation="MELTELLVYMLLTARLTSPAPPACDLRVLSKLLRSHVLA				
	SRISQCEVHPILPTPVLLPAVDPSIGEMKTOMEETKADILGAVTLLEGVMARGL				
	GPTCLSELGOLSGVRRLLLGALOSLIGTOLPGGRTTAHDPAAIFLSFOHLRGV				
	RFLMLVGGSTLCVRRAPPTTAVPSRISLVITLNEIPRTSGLETNTASARTGSGT				
	LKWQOGFRAPKIPGLNOTSRLDIPGLNRIHELKGLPFGPSRRLTGADPIS				
	GTSDTGLSPNLQPGYSPSPHTPGGYTLFPPTLPPTVQHLPLLPDSAPTTP				
	TSPLNTSYTHSONLSQEG"				
	3546. .3673				

exon	/number=3 3960. .4046				
exon	/number=4 5979. .6146				
exon	/number=5 6383. .7576				
exon	/number=6 /number=6				
BASE COUNT	1897 a	2154 c	1807 g	1808 t	
ORIGIN					
Query Match	99.2%;	Score 118;	DB 26;	Length 7666;	
Best Local Similarity	99.2%;	Pred. No. 5.03e-74;			
Matches 118;	Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;
Db	7508	GATCTTTACTCTTGAGAAATGATAGCTTTCTCTCTAGAAATGCTGCTCCCTATACACTAG	7567		
Oy	1	GATCTTTACTCTTGAGAAATGATAGCTTTCTCTCTAGAAATGCTGCTCCCTATACACTAG	60		
Db	7568	ACAAACGAGCCCTGTATAGGAATTAATGGAGCGCCGAAAAAGCTCCCTAAAAAGCAA	7626		
Oy	61	ACAAACGAGCCCTGTATAGGAATTAATGGAGCGCCGAAAAAGCTCCCTAAAAAGCAA	119		
RESULT	3	standard; RNA; HUM; 1721 BP.			
ID	E11961.				
AC	E11961.				
NI	d1110298				
DI	08-OCT-1997 (Rel. 52, Created)				
DT	08-OCT-1997 (Rel. 52, Last updated, Version 1)				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				
DT	Human cDNA encoding thrombopoietin.				
DE	Human cDNA encoding thrombopoietin.				
DR	Human cDNA encoding thrombopoietin.				
DM	Human cDNA encoding thrombopoietin.				

FT source 1. 1721
 FT /organism="Homo sapiens"
 SO Sequence 1721 bp; 413 A; 542 C; 371 G; 395 T; 0 other;

Query Match 58.8%; Score 70; DB 10; Length 1721;
 Best Local Similarity 100.0%; Pred. No. 5,56e-35;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1633 GATCTTTACTCTTGAGAAATGATAGCTTCTCTCAGAAATGCTGTCCCTATACACTAG 1682
 |||||||
 Qy 1 GATCTTTACTCTTGAGAAATGATAGCTTCTCTCAGAAATGCTGTCCCTATACACTAG 60
 |||||||
 Db 1683 ACAAACTGA 1692
 |||||||
 Qy 61 ACAAACTGA 70

RESULT 4 D32047 1691 bp mRNA PRI 28-AUG-1997
 LOCUS Homo sapiens mRNA for thrombopoietin, complete cds.
 DEFINITION D32047
 ACCESSION 92351117
 NID thrombopoietin.
 KEYWORDS Homo sapiens liver cDNA to mRNA.
 SOURCE Homo sapiens
 ORGANISM Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 1691)
 AUTHORS Ogami, K.
 TITLE Submitted (08-JUL-1994) to the DBJ/EMBL/GenBank databases. Kinya
 JOURNAL Ogasmi, KIRIN Brewery Co., Ltd., Pharmaceutical Research Laboratory;
 2-2 Soujiamachi 1 chome, Maebashi, Gunma 371, Japan
 (Tel.:0272-54-8618, Fax:0272-52-2307)
 2 (sites)
 Kato, T., Ogami, K., Shimada, Y., Iwamatsu, A., Sohma, Y., Akahori, H.,
 Horie, K., Kokubo, A., Kudo, Y., Maeda, E., Kobayashi, K., Ohashi, H.,
 Ozawa, T., Inoue, H., Kawamura, K. and Miyazaki, H.
 TITLE Purification and characterization of thrombopoietin
 JOURNAL J. Biochem. 118 (1), 229-236 (1995)
 MEDLINE 96015174
 REMARK Erratum: [[published erratum appears in J Biochem (Tokyo) 1996
 Jan;119(1):208]]

FEATURES
 source
 1. 1691
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="liver"
 sig_peptide
 102..164
 102..1163
 /codon_start=1
 /product="thrombopoietin"
 /db_xref="PID:d1022789"
 /db_xref="PID:g2351118"
 /translation="METELLVLMILTARLTSSPAPACDRLVLSKLRDHSVHL
 SRLQCEVHPPLPTVLLPAVDLSGEKTMETKADILGAVTLLEGVMAARGOL
 GPTCLSLGOLSGVRLALGALSLGTOLPPGRTTAHDPAIFLSFQHLRGV
 RFLMLVGSTLCVRRAPPTTAVPSRTSLVTLNPNSTGLTFNTASARTGSGL
 LKMOGFRAPKIPGLNLSRSLDIPGLNRIHLLNGTRGLFPGSPRRITGAPDIS
 GTSPTGSLPRLQRCYSPSRPHPTGOTTLFPPTLPVYVQLHPLLPDSAPTPTP
 TSPLNSTYTHSONLSQDG"

BASE COUNT 383 a 542 c 371 g 395 t
 ORIGIN

Query Match 58.0%; Score 69; DB 26; Length 1691;
 Best Local Similarity 100.0%; Pred. No. 3.41e-34;
 Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1623 GATCTTTACTCTTGAGAAATGATAGCTTCTCTCAGAAATGCTGTCCCTATACACTAG 1682
 |||||||
 Qy 1 GATCTTTACTCTTGAGAAATGATAGCTTCTCTCAGAAATGCTGTCCCTATACACTAG 60
 |||||||

Db 1683 ACAAACTG 1691
 |||||||
 Qy 61 ACAAACTG 69

RESULT 5 HUMMLCPL 1795 bp mRNA PRI 07-JAN-1995
 LOCUS Human c-mpl ligand (ML) mRNA, complete cds.
 DEFINITION L33410
 ACCESSION 9506826
 NID c-mpl ligand.
 KEYWORDS Homo sapiens cDNA to mRNA.
 SOURCE Homo sapiens
 ORGANISM Eukaryotae; Eukaryotes; Metazoa; Chordata;
 Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 1795)
 de Sauvage, F.J., Haas, P.E., Spencer, S.D., Malloy, B.E., Gurney, R.L.,
 Spencer, S.A., Darbonne, W.C., Hensel, W.J., Wong, S.C., Kuang, W.-J.,
 Oles, K.J., Hultgren, B., Solberg, L.A., Jr., Goeddel, D.V. and
 Eaton, D.L.
 TITLE Stimulation of megakaryocytopoiesis and thrombopoiesis by the c-mpl
 JOURNAL Nature 369 (6481), 533-538 (1994)
 MEDLINE 94261202

FEATURES
 source
 1. 1795
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 sig_peptide
 216..278
 /gene="ML"
 216..1277
 /gene="ML"
 /note="erythropoietin homology domain bp 66..522"
 /codon_start=1
 /product="c-mpl ligand"
 /db_xref="PID:g506827"
 /translation="METELLVLMILTARLTSSPAPACDRLVLSKLRDHSVHL
 SRLQCEVHPPLPTVLLPAVDLSGEKTMETKADILGAVTLLEGVMAARGOL
 GPTCLSLGOLSGVRLALGALSLGTOLPPGRTTAHDPAIFLSFQHLRGV
 RFLMLVGSTLCVRRAPPTTAVPSRTSLVTLNPNSTGLTFNTASARTGSGL
 LKMOGFRAPKIPGLNLSRSLDIPGLNRIHLLNGTRGLFPGSPRRITGAPDIS
 GTSPTGSLPRLQRCYSPSRPHPTGOTTLFPPTLPVYVQLHPLLPDSAPTPTP
 TSPLNSTYTHSONLSQDG"

gene
 mat_peptide
 BASE COUNT 420 a 586 c 382 g 407 t
 ORIGIN

Query Match 34.5%; Score 41; DB 25; Length 1795;
 Best Local Similarity 100.0%; Pred. No. 5.48e-13;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1737 GATCTTTACTCTTGAGAAATGATAGCTTCTCTCAGAAA 1777
 |||||||
 Qy 1 GATCTTTACTCTTGAGAAATGATAGCTTCTCTCAGAAA 41
 |||||||

RESULT 6 I33525 1795 bp DNA PAT 27-JAN-1997
 LOCUS Sequence 1 from patent US 5593666.
 DEFINITION I33525
 ACCESSION 91824316
 NID
 KEYWORDS Unknown.
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1795)
 AUTHORS McDonald, T.P.
 TITLE Methods and compositions for treating thrombocytopenia
 JOURNAL Patent: US 5593666-A 1 14-JAN-1997;

Not pin

```
FEATURES
  source          Location/Qualifiers
                    1..1795
                    /organism="unknown"
BASE COUNT      420 a      586 c      382 g      407 t
ORIGIN
Query Match      34.5%; Score 41; DB 21; Length 1795;
Best Local Similarity 100.0%; Pred. No. 5,48e-13;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1737 GATCTTACTCTTGAGAAATGAAATAGCTTTCTTCAGAAA 1777
0Y 1 GATCTTACTCTTGAGAAATGAAATAGCTTTCTTCAGAAA 41

RESULT 7
LOCUS      E12183      4506 bp      DNA      PAT      27-APR-1998
DEFINITION Human thrombopoietin gene.
ACCESSION E12183
NID        93251017
KEYWORDS   JP 1996277296-A/3.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens.
REFERENCE   Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Homiidae; Homo.
            Miyazaki, H., Kato, H., Ogami, K., Iwamatsu, A., Akahori, H., Kuroki, R.,
            Shimizu, T. and Mito, T.,
            1 (bases 1 to 4506)
            PROTEIN HAVING TPO ACTIVITY
            Patent: JP 1996277296-A 3 22-OCT-1996;
            KIRIN BREWERY CO LTD
COMMENT     OS Homo sapiens (human)
            PN JP 1996277296-A/3
            PD 22-OCT-1996
            PE 14-FEB-1995 JP 94P 1995061363
            PR 14-FEB-1994 JP 94P 39090, 25-MAR-1994 JP 94P 79842, PR
            01-JUN-1994 JP 94P 155126, 15-JUN-1994 JP 94P 167328, PR
            17-AUG-1994 JP 94P 227159, 17-AUG-1994 JP 94P 193169, PR
            01-NOV-1994 JP 94P 304167, 01-DEC-1994 JP 94P 298669, PR
            28-DEC-1994 JP 94P 341200
            PI MIYAZAKI HIROSHI, KATO HISASHI, OGAMI KINYA, IWAMATSU AKIHIKO,
            AKAHORI HIROKORI, KUROKI RYOTA, SHIMIZU TOSHIYUKI, PI MITO
            TAKANORI
            PC C07K14/52, C07H21/04, C12N15/09, C12P21/02//A61K38/00, (C12P21/02,
            PC C12R1:19),
            PC (C12P21/02, C12R1:91);
            CC strandedness: Double;
            CC topology: Linear;
            FH key
            FH source
            FT 1..4506
            FT /organism="Homo sapiens"
            FT intron 284..514
            FT exon 545..642
            FT intron 643..928
            FT exon 929..1015
            FT intron 1016..2947
            FT exon 2948..3115
            FT intron 3116..3351
            FT CDS
            FT join(271..283,515..642,929..1015,2948..3115,3352..4014) FT
            FT sig_peptide join(271..283,515..564).
            FT location/Qualifiers
            FT 1..4506
            FT /organism="Homo sapiens"
            FT /db_xref="taxon:9606"
BASE COUNT      1149 a      1277 c      972 g      1108 t
ORIGIN
Query Match      25.2%; Score 30; DB 21; Length 4506;
Best Local Similarity 100.0%; Pred. No. 1.97e-05;
```

```
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 4477 GATCTTACTCTTGAGAAATGAAATAGCTT 4506
0Y 1 GATCTTACTCTTGAGAAATGAAATAGCTT 30

RESULT 8
ID      E11962      standard; DNA; HUM; 4506 BP.
AC      E11962;
NT      d1110299
DT      08-OCT-1997 (Rel. 52, Created)
DE      08-OCT-1997 (Rel. 52, Last updated, Version 1)
DE      Human genomic DNA sequence encoding thrombopoietin.
KW      JP 1996228781-A/3.
OS      Homo sapiens (human)
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates;
OC      Catarrhini; Homiidae; Homo.
RN      [1]
RP      1-4506
RA      Miyazaki H., Kato H., Ogami K., Iwamatsu A., Akahori H., Kuroki R.,
RA      Shimizu T., Mito T.;
RT      "DNA CODING FOR PROTEIN HAVING TPO ACTIVITY";
RL      Patent number JP 1996228781-A/3, 10-SEP-1996.
RL      KIRIN BREWERY CO LTD.
OS      Homo sapiens (human)
CC      PN JP 1996228781-A/3
CC      PD 10-SEP-1996
CC      PE 14-FEB-1995 JP 94P 1995063298
CC      PR 14-FEB-1994 JP 94P 39090, 25-MAR-1994 JP 94P 79842,
CC      PR 01-JUN-1994 JP 94P 155126, 15-JUN-1994 JP 94P 167328,
CC      PR 17-AUG-1994 JP 94P 227159, 17-AUG-1994 JP 94P 193169,
CC      PR 01-NOV-1994 JP 94P 304167, 01-DEC-1994 JP 94P 298669,
CC      PR 28-DEC-1994 JP 94P 341200
CC      PI MIYAZAKI HIROSHI, KATO HISASHI, OGAMI KINYA, IWAMATSU AKIHIKO,
CC      PI AKAHORI HIROKORI, KUROKI RYOTA, SHIMIZU TOSHIYUKI,
CC      PI MITO TAKANORI
CC      PC C12N15/09, A61K38/00, C07K14/52, C12N1/21, C12N5/10, C12P21/02,
CC      PC (C12N1:21),
CC      PC (C12N1:19), (C12N5/10, C12R1:91), (C12P21/02, C12R1:19), (C12P21/02,
CC      PC C12R1:91);
CC      CC strandedness: Double;
CC      CC topology: Linear;
CC      FH key
CC      FH source
CC      FT 1..4506
CC      FT /organism="Homo sapiens"
CC      FT intron 284..514
CC      FT exon 515..642
CC      FT intron 643..928
CC      FT exon 929..1015
CC      FT intron 1016..2947
CC      FT exon 2948..3115
CC      FT intron 3116..3351
CC      FT CDS
CC      FT join(271..283,515..642,929..1015,2948..3115,3352..4014)
CC      FT 2..4014)
CC      FT /product="thrombopoietin precursor"
CC      FT join(271..283,515..564)
CC      FT sig_peptide
CC      FT mat_peptide
CC      FT join(565..642,929..1015,2948..3115,3352..4014)
CC      FT /product="thrombopoietin"
CC      FT replace(253,"a")
CC      FT variation replace(3584,"a")
CC      FT variation replace(4052,"g")
CC      FT key
FH      FH location/Qualifiers
FH      FT source
FH      FT 1..4506
FH      FT /organism="Homo sapiens"
FH      FT Sequence 4506 BP; 1149 A; 1277 C; 972 G; 1108 T; 0 other;
SQ
Query Match      25.2%; Score 30; DB 10; Length 4506;
Best Local Similarity 100.0%; Pred. No. 1.97e-05;
```

Matches	30;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Db	4477	GATCTTACTCTTGAGAAATGAATAAGCTT	4506						
0y	1	GATCTTACTCTTGAGAAATGAATAAGCTT	30						
RESULT	9	E12215	4506 bp	DNA	PAT	27-APR-1998			
LOCUS									
DEFINITION									
ACCESSION		E12215							
NID		93251049							
KEYWORDS		JP 1996291196-A/2.							
SOURCE									
ORGANISM		Homo sapiens.							
REFERENCE		Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.							
AUTHORS		1 (bases 1 to 4506)							
TITLE		Miyazaki, H., Kato, H., Ogami, K., Iwamatsu, A., Akahori, H., Kuroki, R., Shimizu, T., and Muto, T.							
JOURNAL		PROTEIN HAVING TPO ACTIVITY							
COMMENT		Patent: JP 1996291196-A 2 05-NOV-1996;							
		KIRIN BREWERY CO LTD							
		OS Homo sapiens (human)							
		PN JP 1996291196-A/2							
		PD 05-NOV-1996							
		PR 14-FEB-1995 JP 1995355052							
		PR 14-FEB-1994 JP 94P 39090, 25-MAR-1994 JP 94P 79842, PR							
		01-JUN-1994 JP 94P 155126, 15-JUN-1994 JP 94P 167328, PR							
		17-AUG-1994 JP 94P 227159, 17-AUG-1994 JP 94P 193169, PR							
		01-NOV-1994 JP 94P 304167, 01-DEC-1994 JP 94P 298669, PR							
		28-DEC-1994 JP 94P 341200							
		PI MIYAZAKI HIROSHI, KATO HISASHI, OGAMI KINYA, IWAMATSU AKIHITO,							
		PI AKAHORI HIROHORI, KUROKI RYOTA, SHIMIZU TOSHIYUKI, PI MUTO							
		TAKANORI							
		PC C07K14/53.C07K1/22.C12N15/09//A61K38/00.C12P21/08; CC							
		strandedness: Double;							
		CC topology: Linear;							
		FH key							
		FH Location/Qualifiers							
		FT source							
		FT 1..4506							
		FT /organism="Homo sapiens"							
		FT 284..514							
		FT intron							
		FT exon							
		FT 515..642							
		FT intron							
		FT exon							
		FT 643..928							
		FT exon							
		FT 929..1015							
		FT intron							
		FT 1016..2847							
		FT exon							
		FT 2848..3115							
		FT intron							
		FT 3116..3351							
		FT CDS							
		join(271..283,515..642,929..1015,2948..3115,3352..4017) FT							
FEATURES		/product="thrombopoietin";							
Source		FT sig_peptide							
		Location/Qualifiers							
		1..4506							
		/organism="Homo sapiens"							
		/db_xref="taxon:9606"							
BASE COUNT		1149 a 1277 c 972 g 1108 t							
ORIGIN									
Query Match		25.2%;							
Best Local Similarity		100.0%;							
Matches		30; Conservative							
		0; Mismatches							
		0; Indels							
		0; Gaps							
Db	4477	GATCTTACTCTTGAGAAATGAATAAGCTT	4506						

ACCESSION	166494
NID	92724471
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified. 1 (bases 1 to 7218)
AUTHORS	Dorner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE	Recombinant fowlpox virus
JOURNAL	Patent: US 5670367-A 14 23-SEP-1997;
FEATURES	Location/Qualifiers 1..7218
SOURCE	1..7218
BASE COUNT	1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN	
Query Match	21.0% Score 25; DB 21; Length 7218;
Best Local Similarity	4.4% Pzed. No. 2.79e-02;
Matches	3; Conservative 43; Mismatches 22; Indels 0; Gaps 0;
Db	1370 yyy 1429
Cp	120 TTTCCTTTTANAGACCTTTGGCGCCCTTAATTCCTATACAGCCACATTTGT 61
Db	1430 yyyyyyGA 1437
Cp	60 CTAGTGA 53
RESULT	11
LOCUS	CEP1A1 39353 bp DNA INV 21-SEP-1998
DEFINITION	Caenorhabditis elegans cosmid F1A1, complete sequence.
ACCESSION	250857
NID	9567024
KEYWORDS	HTG: Nuclear hormone receptor; Transfer-RNA; tRNA-glu.
SOURCE	Caenorhabditis elegans.
ORGANISM	Caenorhabditis elegans
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis 1 (bases 1 to 39353)
AUTHORS	Wilson, R., Alnscough, R., Anderson, K., Baynes, C., Betks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Lathille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roop, A., Saunders, D., Showkeen, R., Smaldon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkenson-Sprout, J. and Mohlman, P.
TITLE	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL	Nature 368 (6465), 32-38 (1994)
MEDLINE	94150718
REFERENCE	2 (bases 1 to 39353)
AUTHORS	Harris, B.
JOURNAL	Direct Submission
COMMENT	Submitted (06-NOV-1996) Louis, MO 63110, USA. E-mail: jess@sanger.ac.uk or tw@nemata.de.wustl.edu Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information. For a graphical representation of this sequence and its analysis see: http://webc.sanger.ac.uk/cgi-bin/display?db=wormace&class=sequence&object=F1A1 Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is NOT necessarily the entire insert of

the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. This sequence is the entire insert of clone F11A1. The true right end of clone M79 is at 5947 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence 250806. The end of this sequence (39254..39353) overlaps with the start of sequence AL009066.

```

FEATURES
source
1..39353
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="X"
/clone="F11A1"
17129..17456
/gene="F11A1.1"
join(17129..17302,17352..17456)
/gene="F11A1.1"
/codon_start=1
/db_xref="PID:e348773"
/db_xref="PID:g967025"
/db_xref="SPTREMBL:Q19331"
/translation="MYSNFSHNRKLPYPRPYFSAVSYETPIQTSKTPGSASIVD
EKIDDDKSSPASRRHRHSKNCLTPPNSDDVSLLCFENOVRL"
complement(28112..31091)
/gene="F11A1.2"
complement(join(28112..28424,28478..28518,29549..29817,
31055..31091))
/gene="F11A1.2"
/codon_start=1
/db_xref="PID:e349167"
/db_xref="PID:g967026"
/db_xref="SPTREMBL:Q19332"
/translation="MYNPOINSECAKKKSKAVRRKNEKMGKTGGRVVNCI
LNRVCVCUKPLTSFARVQNEFCFHSIRRDPADETHLSPETKRRKVLNSHNSMT
PRRATGCVNGWRCPTFSIVRVATLKSSEVRNEKRLALRLAGVSELOTFEPVYL
ASIVSWGTALSHLSTLSNLVQFTLRALITKTVKLIVEAREKCSVEFFNPLPQIK
L"
31980..36995
/gene="F11A1.3"
join(31980..32171,32311..32374,32817..32899,32947..33069,
33115..33381,33448..33495,33544..33615,34309..34425,
34764..34838,35174..35271,35321..35672,35920..36076,
36160..36436,36484..36645,36692..36767,36903..36995)
/gene="F11A1.3"
/note="similar to nuclear hormone receptor"
/codon_start=1
/db_xref="PID:e348774"
/db_xref="PID:g967027"
/db_xref="SPTREMBL:Q19333"
/translation="MFHQQPPVVDLNNLFMYHLVSEFNTYQCQLQIHNYSPYLT
POIKCEQKRWHRKIKSPOMSTPDDPIYHSPISGAPGVHSGSVLKSEESGACG
SPDGLDSSSESRROKTCRGDHAFTGNNFTQESCAPFRNALPKRCKCPY
SECEITSYVRRROCKRLKCTVGMKKEMLINEEDLRNRKRLNNTGCKNSRSP
GNOCSPGQWQHLSFHHGVAIYPPORPRLTINMDNMHHMANPBNAPOLI
SPFOADPYPLTSPVSSADSPPNRSULTMHNKSGSDGVSAGXKDYINMRAAP
SFFNRRKMSGVVLTSTEEKQLSRIPGAVGLNDEEPIKKRAIYONGPMFA
ETTPYSAEMKSLSRHNSSTSGTEKNHMTSVSAIPGSAQHEDIASFGGIVT
ATGGGAAEEYKRMNMEYENCIOALSADSPENCPOEAMIPEEYMTPHGVOYS
DPVOVPAERININOLNAAELKALDAVEARYGMDPMEOGPDIMINIDVTRFVY
AKGPAFREVSQEGKFSLLKGMTEMLTVRGVTYDASTNEFKPTLKGQVSVNVD
MFAKLNNAAQOKAKCLEFFGFEDEETIKKNLAIYLVMLAVLSVRSPPNENDVRI
VTERHNHFMSTLNYLLESLFGEQARRIFERIPKALGLNLETARNAGMLFMGTVSGEA
EELPGEFFIK"
complement(38536..38607)
/gene="F11A1.t2"
/note="TTC Glu E-tRNA: predicted using tRNAscan-SE-1.11;
preliminary prediction, similar to tRNA-Glu"
complement(38536..38607)
/gene="F11A1.t2"
39107..39178
/gene="F11A1.t1"
/note="TTC Glu E-tRNA: predicted using tRNAscan-SE-1.11;
preliminary prediction, similar to tRNA-Glu"

```

```

gene
39107..39178
/gene="F11A1.t1"
join(complement(380..503),complement(234..321),
complement(73..184),complement(250806:33620..33673),
complement(250806:33444..33565),
complement(250806:33155..33242),
complement(250806:33155..33242),
complement(250806:32838..33101)
/gene="M79.2"
join(complement(380..503),complement(234..321),
complement(73..184),complement(250806:33620..33673),
complement(250806:33444..33565),
complement(250806:33155..33242),
complement(250806:32838..33101)
/gene="M79.2"
/codon_start=1
/db_xref="PID:e1322086"
/db_xref="PID:g3642276"
/translation="MCHIVPVVYLRGGQPSDVWLKLCFFPLASFVFLFKYML
LSIFYGSCITLTFSTVLIFGLSGVLPQWTLHYVAMIOGFFPREKLENTYNE
ESEQRSEKSDQIAERVKRIIDIDGPGFOAPTLRPTLGMPEKNNSTERKAT
FAGLIFRIFFLMPVRCILTSFVVALAGLQTAFTRLSDREKTVAIVCRFGSM
GLVANYRNPQRPKKPGVAVSNHLPNDIQILMAGTPhgssgyvvtgQKHKGIIGIY
CYLRVVC"
BASE COUNT      12348 a      7137 c      6554 g      13314 t
ORIGIN
Query Match      19.3%; Score 23; DB 18; Length 39353;
Best Local Similarity 92.3%; Pred. No. 4.29e-01;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 26783 TTTGCTTTTAGGACCTTTTAGCG 26808
CP 120 TTTGCTTTTAGGACCTTTTAGCG 95
RESULT 12
LOCUS HSEC200CB 5975 bp DNA PRI 17-NOV-1995
DEFINITION Human scRNA (BC200 beta) pseudogene.
ACCESSION U01307
NID 9401721
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 5975)
Martignetti,J.A. and Brosius,J.
BC200 RNA: a neural RNA polymerase III product encoded by a
monomeric Alu element
Proc. Natl. Acad. Sci. U.S.A. 90 (24), 11563-11567 (1993)
JOURNAL 94089673
MEDLINE 2 (bases 1 to 5975)
REFERENCE Tiedge,H., Chen,W. and Brosius,J.
AUTHORS Primary structure, neural-specific expression, and dendritic
TITLE location of human BC200 RNA
JOURNAL J. Neurosci. 13 (6), 2382-2390 (1993)
MEDLINE 93274370
REFERENCE 3 (bases 1 to 5975)
AUTHORS Brosius,J.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-1993) Juergen Brosius, Fishberg Research Center
f. Neurobiology, Mount Sinai School of Medicine, One Gustave L.
Levy Place, New York, NY 10029, USA
FEATURES
source
1..5975
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="adult"
1902..1914
/rpt_type="flanking"
1915..2219
/rpt_family="Alu, Sx subfamily [Jurka and Milosavljevic

```

(1991) J. Mol. Evol. 32, 105-131]."

repeat_region 2220..2232
/rpt-type=flanking

repeat_region 2894..2905
/rpt-type=flanking

repeat_unit complement(2906..3203)
/rpt-family="Alu, Sg subfamily [Jurka and Milosavljevic,
Op. cit.]"]

repeat_region 3204..3215
/rpt-type=flanking

repeat_region 4644..4657
/rpt-type=flanking

repeat_unit 4658..5041
/note="Orthologous locus of BC200 beta pseudogene in chimpanzee and other apes has a normal Alu Sb dimer. Replacement of the left Alu Sb monomer with a BC200 pseudogene occurred <6 MYR."

repeat_unit 4658..5041
/note="While the true gene only involves a monomeric Alu element, FUM-CL, the pseudogene is a composite dimer: BC200-like as left arm and a member of the Alu Sb subfamily as right arm."

scRNA 4658..4860
/gene="BC200 beta"
/note="Only three changes compared to the true gene BC200 alpha. Thus far, however, no transcripts detected that would indicate that the BC200 beta pseudogene is transcriptionally active."
/pseudo

gene 4658..4860
/gene="BC200 beta"

repeat_region 5042..5055
/rpt-type=flanking

BASE COUNT 1768 a 1256 c 1416 g 1535 t

ORIGIN

Query Match 18.5%; Score 22; DB 25; Length 5975;
Best Local Similarity 80.6%; Prid. No. 1.61e+00;
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 3065 GGCTAATTTTGTATGATAGACAGACGCAATTTC 3100
||||| ||||| ||||| ||| ||||| |||||
Cp 73 GGCTCAGTTTGCTAGTGTATAGGACAGCATTTTC 38

RESULT 13

LOCUS AC004341 61964 bp DNA INV 28-JUL-1998

DEFINITION Drosophila melanogaster DNA sequence (PI DS02473 (D239)), complete sequence.

ACCESSION AC004341

NID 93342724

KEYWORDS HTG.

SOURCE Drosophila melanogaster (Subclones in sac from P1 clone DS02473 (D239)) DNA.

ORGANISM Drosophila melanogaster

AUTHORS Eukariyola, Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachytera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 61964)

REFERENCE Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A., Svitskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E., Blazer,J.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazza,P., Mok,M.S., Mosherfi,A.R., Moshrefi,M., Nixon,K.K., Paclio,J.M., Park,S., Pfeiffer,B., Punch,E., Snit,E., Twomey,B., Wan,K.H., Whitefaw,K.R., Yee,A., Zhang,R., Zieren,L.L. and Kimmel,B.
Sequencing of Drosophila chromosome 2R, region 53D2-53D10 Unpublished (1997)
2 (bases 1 to 61964)

TITLE CeInker,S.E., George,R.A., Galle,R., Svitskas,R.R., Hoskins,R.A., Agabayani,A., Arcaina,T.T., Baxter,E., Blazer,J.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,

REFERENCE AUTHORS

TITLE	JOURNAL	Hummetti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B., Lomoti,M.A., Mak,J., Mazda,P., Mok,W.S., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacled,J.M., Park,S., Pfeiffer,B., Punche,E., Sitr,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.E.
TITLE	JOURNAL	Submitted (07-MAR-1998) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT		On Jul 28, 1998 this sequence version replaced g1:3334954. Sequence submitted by: Lawrence Drosophila Genome Project Berkeley Drosophila National Laboratory, MS 64-121 Berkeley, CA 94720
FEATURES	source	For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://fruitfly.berkeley.edu/sequence/) or send email to drosophila@mbc.lbl.gov . Library location: 73-26.
BASE COUNT	ORIGIN	1. 61964 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /chromosome="2R" /map="53D2-53D10" /clone="P1 DS02473 (D239)" /note="Between bp 17,024 and bp 17,025 ISI element sequence deleted."
Query Match		18.5%; Score 22; DB 18; Length 61964;
Best Local Similarity		66.2%; Pred. No. 1.61e+00;
Matches		45; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
Db	34163	ATATTTCTATTTCGTATGATGATGATTTCTTTAAACAACATCTAGTCATATAAATAAA 34222
Qy	2	ATCTTACTCTTGAGAAATGATAAGCTTTCTCTCAGAAATGCTGCTCATACACTGA 61
Db	34223	CAAACTG 34230
Qy	62	CAAACTG 69
RESULT	14	
LOCUS	AC005369	74371 bp DNA PRI 01-AUG-1998
DEFINITION	Homo sapiens chromosome 5, BAC clone 119j3 (BRL H175), complete sequence.	
ACCESSION	AC005369	
NID	93367505	
KEYWORDS	HTG.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
AUTHORS	1 (bases 1 to 74371) Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M., Kader,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M., Rojeski,H., Subramanian,S. and Martin,C.H.	
TITLE	JOURNAL	Sequencing of human chromosome 5
REFERENCE	JOURNAL	Unpublished
AUTHORS		2 (bases 1 to 74371)
TITLE		Ricke,D.O.
JOURNAL		Large Scale Sequence Analysis and Annotation with the Sequence Comparison Analysis (SCAN) System
REFERENCE		Unpublished
AUTHORS		3 (bases 1 to 74371) Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M., Davis,C.A., Kader,K., Miguel,T., Pitluck,S., Pollard,M., Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE	JOURNAL	Direct Submission
JOURNAL		Submitted (10-AUG-1998) Human Genome Center, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, MS 74-157, Berkeley, CA 94720, U.S.A.

```
COMMENT      Sequence submitted by:
FEATURES     DOE Joint Genome Institute.
SOURCE       Location/Qualifiers
            1. 74371
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /map="5q"
               /clone="119j3"
               /chromosome="5"
               /note="LBNL H175"
repeat_region /rpt_family="Alu"
               893. .1030
repeat_region /rpt_family="Alu"
               2285. .2438
               /rpt_family="Alu"
               2818. .2859
               /note="(Gr)21"
               /rpt_type=tandem
               /rpt_unit=GT
misc_feature  join(3246..3410,3721..3828)
               /standard_name="ALF"
               /note="658 & 698 protein identity GenPept:U2377"
repeat_region /rpt_family="Alu"
               3431. .3724
               /rpt_family="Alu"
               3707. .3728
               /note="(A)22"
               /rpt_type=tandem
               /rpt_unit=A
repeat_region /rpt_family="Alu"
               4366. .4661
repeat_region /rpt_family="Alu"
               5327. .5602
repeat_region /rpt_family="Alu"
               6586. .6956
repeat_region /rpt_family="L1"
               6647. .6684
               /note="(CA)19"
               /rpt_type=tandem
               /rpt_unit=CA
repeat_region /rpt_family="Alu"
               7113. .7373
               /standard_name="possible repeat"
repeat_region /rpt_family="Alu"
               8258. .8503
repeat_region /rpt_family="Alu"
               9070. .9387
repeat_region /rpt_family="Alu"
               complement(9740..9845)
repeat_region /rpt_family="MER42"
               complement(10440..11015)
repeat_region /rpt_family="Alu"
               11950. .12250
repeat_region /rpt_family="Alu"
               12057. .12085
               /note="(A)29"
               /rpt_type=tandem
               /rpt_unit=A
repeat_region /rpt_family="Alu"
               12365. .12645
               /rpt_family="Alu"
               13727. .13750
               /note="(AC)12"
               /rpt_type=tandem
               /rpt_unit=AC
repeat_region /rpt_family="L1"
               13783. .14024
repeat_region /rpt_family="L1"
               14175. .14470
               /rpt_family="Alu"
               complement(14906..15259)
               /standard_name="possible repeat"
repeat_region /rpt_family="Alu"
               15300. .15613
repeat_region /rpt_family="Alu"
               16671. .16690
               /note="(T)20"
               /rpt_type=tandem
               /rpt_unit=T
repeat_region complement(16675..16977)
               /rpt_family="Alu"
repeat_region complement(16993..17085)
               /rpt_family="MER42"
               complement(17678..18276)
               /rpt_family="Alu"
               19305. .19583
repeat_region /rpt_family="Alu"
               19914. .19945
               /note="(GTT)8"
               /rpt_type=tandem
               /rpt_unit=GTT
repeat_region complement(19943..20222)
               /rpt_family="Alu"
               21202. .21496
repeat_region /rpt_family="Alu"
               21736. .22035
repeat_region /rpt_family="Alu"
               22017. .22038
               /note="(A)22"
               /rpt_type=tandem
               /rpt_unit=A
repeat_region /rpt_family="Alu"
               22321. .22457
               /rpt_family="Alu"
               complement(22638..22981)
               /standard_name="possible repeat"
               23022. .23326
repeat_region /rpt_family="Alu"
               23473. .23761
repeat_region /rpt_family="Alu"
               23744. .23767
               /note="(A)24"
               /rpt_type=tandem
               /rpt_unit=A
repeat_region complement(24639..24694)
               /rpt_family="MER42"
               complement(25349..25713)
               /rpt_family="Alu"
               complement(25727..26471)
               /rpt_family="Alu"
               27191. .27477
repeat_region /rpt_family="Alu"
               27774. .28057
repeat_region /rpt_family="Alu"
               28040. .28066
               /note="(A)27"
               /rpt_type=tandem
               /rpt_unit=A
repeat_region /rpt_family="MER20"
               28712. .28930
               /rpt_family="MER20"
               complement(28769..28838)
               /note="GRAIL 2 excellent exon, frame 0"
               28987. .29214
repeat_region /rpt_family="Alu"
               join(29485..29559,29598..29779)
               /note="95% identity dbSTS:G14522 (SHGC-11312)"
               29495. .29976
               /standard_name="A1027942"
               /note="100% identity EST ov84a10.x1"
               complement(30401..30536)
               /note="GRAIL 2 excellent exon, frame 2"
               /rpt_family="MIR"
               complement(31573..31724)
               /note="GRAIL 2 excellent exon, frame 1"
               complement(32159..32332)
               /note="GRAIL 2 excellent exon, frame 2"
               32388. .32488
               /rpt_family="MLT1"
               32617. .32908
               /rpt_family="Alu"
               32977. .33088
               /rpt_family="MLT1"
```

```
repeat_region complement(16993. .17085)
repeat_region /rpt_family="MER42" .18276)
repeat_region complement(17678 .18276)
repeat_region /rpt_family="Alu"
repeat_region 19305. .19583
repeat_region /rpt_family="Alu"
repeat_region 19914. .19945
repeat_region /note="(GTTT)8"
repeat_region /rpt_type=tandem
repeat_region /rpt_unit=Grif
repeat_region complement(19943. .20222)
repeat_region /rpt_family="Alu"
repeat_region 21202. .21496
repeat_region /rpt_family="Alu"
repeat_region 21736 .22035
repeat_region /rpt_family="Alu"
repeat_region 22017. .22038
repeat_region /note="(A)22"
repeat_region /rpt_type=tandem
repeat_region /rpt_unit=A
repeat_region 22321. .22457
repeat_region /rpt_family="Alu"
repeat_region complement(22638. .22981)
repeat_region /standard_name="possible repeat"
repeat_region 23022. .23326
repeat_region /rpt_family="Alu"
repeat_region 23473. .23761
repeat_region /rpt_family="Alu"
repeat_region 23744. .23767
repeat_region /note="(A)24"
repeat_region /rpt_type=tandem
repeat_region /rpt_unit=A
repeat_region complement(24639. .24694)
repeat_region /rpt_family="MER42" .25713)
repeat_region complement(25349. .25713)
repeat_region /rpt_family="Alu"
repeat_region complement(25727. .26471)
repeat_region /rpt_family="Alu"
repeat_region 27191. .27477
repeat_region /rpt_family="Alu"
repeat_region 27774. .28057
repeat_region /rpt_family="Alu"
repeat_region 28040. .28066
repeat_region /note="(A)27"
repeat_region /rpt_type=tandem
repeat_region /rpt_unit=A
repeat_region 28712. .28930
repeat_region /rpt_family="MER20"
repeat_region complement(28769. .28838)
repeat_region /note="GRAIL 2 excellent exon, frame 0"
repeat_region 28987. .29214
repeat_region /rpt_family="Alu"
repeat_region join(29485. .29559,29598. .29779)
repeat_region /note="95% identity dbSTS:G14522 (SHGC-11312)"
repeat_region 29495. .29976
repeat_region /standard_name="AL027942"
repeat_region /note="100% identity EST ov84a10.xl"
repeat_region complement(30401. .30536)
repeat_region /note="GRAIL 2 excellent exon, frame 2"
repeat_region complement(30682. .30733)
repeat_region /rpt_family="MIR"
repeat_region complement(31573. .31724)
repeat_region /note="GRAIL 2 excellent exon, frame 1"
repeat_region complement(32159. .32232)
repeat_region /note="GRAIL 2 excellent exon, frame 2"
repeat_region 32388. .32488
repeat_region /rpt_family="MLT1"
repeat_region 32617. .32908
repeat_region /rpt_family="Alu"
repeat_region 32977. .33088
repeat_region /rpt_family="MLT1"
repeat_region complement(33670. .33785)
repeat_region /rpt_family="Alu"
```

```
misc_feature complement(34021. .34144)
misc_feature /note="GRAIL 2 excellent exon, frame 2"
misc_feature complement(35238. .35331)
misc_feature /note="GRAIL 2 excellent exon, frame 1"
repeat_region 36392. .36663
repeat_region /rpt_family="Alu"
misc_feature 36901. .37222
misc_feature /note="100% identity EST ou55c09.xl"
STS /db_xref="dbEST:AI025011"
STS 36901. .37164
STS /standard_name="STSG-9983"
STS /db_xref="dbSTS:G26854"
STS complement(37595. .37654)
misc_feature /note="GRAIL 2 excellent exon, frame 0"
misc_feature complement(join(38063. .38218,38462. .38578,38741. .38995,
39071. .39205,39532. .39630,39935. .40048,40300. .40410,
40503. .40661,41868. .41972,42103. .42225,42492. .42569,
44379. .44507))
misc_feature /standard_name="histidyl-tRNA synthetase"
misc_feature /note="78%-100% protein identity GenPept.U18937"
misc_feature complement(38069. .38215)
misc_feature /note="GRAIL 2 excellent exon, frame 0"
misc_feature complement(38462. .38578)
```

...: remainder of annotations omitted.

```
Query Match 18.5%; Score 22; DB 26; Length 74371;
Best Local Similarity 16.5%; Pred.No.1.61e+00;
Matches 17; Conservative 48; Mismatches 37; Indels 1; Gaps 1;

Db 15975 SVCYCSGKYYWCRCMYCYYSYKYWMSYCTCTSWGHRWWSKGRSMWYASRSCG 16034
Cp 104 CTTTCGGGGCTC-CCATTATTCTCTATACAGGCTCAGTTTGTCTAGTGTATAGGGAC 46
Db 16035 SCSCSMCMCMRCSCMSMKMWWTTTTTTTTRTWTWTKWKAGA 16077
Cp 45 AGCATTTCTGAGAGAAAGCTATTTCATTTCTCAAGAGTAAGA 3
```

Search completed: Tue Dec 22 05:08:46 1998
Job time : 236 secs.

FEATURES		Location/Qualifiers	
Source		1. .74371	
repeat_region	complement(33670. .33785)	/organism="Homo sapiens"	
misc_feature	/rpt_family="Alu"	/db_xref="taxon:9606"	
misc_feature	/note="GRAIL 2 excellent exon, frame 2"	/map="5q"	
repeat_region	complement(35238. .35331)	/clone="119j3"	
misc_feature	/note="GRAIL 2 excellent exon, frame 1"	/chromosome="5"	
repeat_region	36392. .36663	/note="LBNL H175"	
misc_feature	/rpt_family="Alu"	893. .1030	
STS	35901. .37222	/rpt_family="Alu"	
misc_feature	/note="100% identity EST ou55c09.x1"	2295. .2438	
		/db_xref="dbEST:AI025011"	
misc_feature	/db_xref="dbSTS:G26554"	2618. .2859	
		/note="(GT)21"	
misc_feature	complement(37595. .37654)	/rpt_type=tandem	
		/rpt_unit=GT	
misc_feature	/note="GRAIL 2 excellent exon, frame 0"	join(3246. .3410,3721. .3828)	
		/standard_name="RLF"	
misc_feature	39071. .39205,39532. .39630,39935. .40048,40300. .40410,40503. .40661,41868. .41972,42103. .42225,42492. .42569,44379. .44507)	/note="65% & 69% protein identity GenPept:U22377"	
		3431. .3724	
misc_feature	/standard_name="histidyl-tRNA synthetase"	/rpt_family="Alu"	
		3707. .3728	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/note="(A)22"	
		/rpt_type=tandem	
misc_feature	/note="GRAIL 2 excellent exon, frame 0"	/rpt_unit=A	
		4366. .4661	
misc_feature	complement(38462. .38578)	/rpt_family="Alu"	
		5327. .5602	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_family="Alu"	
		6586. .6956	
misc_feature	/note="GRAIL 2 excellent exon, frame 0"	/rpt_family="L1"	
		6647. .6684	
misc_feature	complement(38462. .38578)	/note="(CA)19"	
		/rpt_type=tandem	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_unit=CA	
		7113. .7373	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_family="Alu"	
		complement(7830. .8185)	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/standard_name="possible repeat"	
		8258. .8503	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_family="Alu"	
		9070. .9387	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_family="Alu"	
		complement(9740. .9845)	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_family="MER42"	
		complement(10440. .11015)	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_family="Alu"	
		11950. .12250	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_family="Alu"	
		12057. .12085	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/note="(A)29"	
		/rpt_type=tandem	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_unit=A	
		12365. .12645	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_family="Alu"	
		13727. .13750	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/note="(AC)12"	
		/rpt_type=tandem	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_unit=AC	
		13783. .14024	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_family="L1"	
		14175. .14470	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_family="Alu"	
		complement(14906. .15259)	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/standard_name="possible repeat"	
		15300. .15613	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_family="Alu"	
		16671. .16690	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/note="(T)20"	
		/rpt_type=tandem	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_unit=T	
		complement(16675. .16977)	
misc_feature	/note="78% 100% protein identity GenPept:U18937"	/rpt_family="Alu"	